My Scrawl Hides It All: Protecting Text Messages Against Shoulder Surfing With Handwritten Fonts

Malin Eiband, Emanuel von Zezschwitz, Daniel Buschek, Heinrich Hußmann

University of Munich (LMU), Germany
{malin.eiband, emanuel.von.zezschwitz, daniel.buschek,
heinrich.hussmann}@ifi.lmu.de

Abstract

We present a novel concept for protecting text messages (e.g. notifications) on mobile devices from shoulder surfing. We propose to display the text in the user's handwriting, assuming that people can read their own handwriting easier and faster than strangers. Our approach was evaluated in a proof-of-concept user study that revealed significant differences in reading time: Participants were indeed slower when reading the unfamiliar handwriting of the other participants compared to their own, and they tended to make more errors. Even though this effect was not present for all participants, we argue that our results may provide the basis for protection mechanisms applicable in real-world scenarios.

Author Keywords

Visual Privacy; Shoulder Surfing; Observability; Mobile Device; Handwriting; Usable Security.

ACM Classification Keywords

H.5.2 [Screen design (e.g., text, graphics, color)]: User Interfaces

Introduction

Mobile devices (e.g. smartphones) are frequently used in public settings like a bus, the subway or a crowded cafe, which might expose sensitive content to bystanders. The

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
Copyright © ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892511

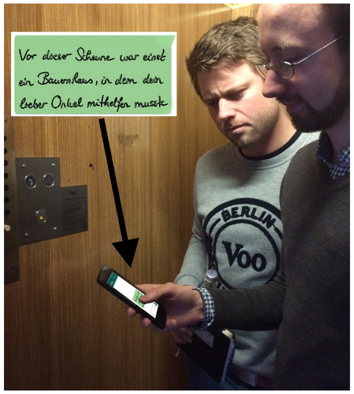

Vor dieser Scheune war einst
ein Bauernhaus, in dem dein
lieber Onkel mithelfen musste

Figure 1: A user is reading text messages on his smartphone. The print font of his device is replaced with his handwriting, so that he can easily read the message while the shoulder surfer next to him has difficulties deciphering the unfamiliar font.

act of observing other people's information without their permission is commonly called *shoulder surfing* and has been intensively studied. To date however, much research has focused on protecting password or PIN entry from shoulder surfing (i.a. [8, 16, 17]), although the actual threat in real-world use has been questioned [5, 9]. In contrast, *visual privacy*, that is privacy issues arising from shoulder surfing, also when inadvertent and non-malicious, is comparably underexplored. Related research emphasizes the importance of visual privacy and shows that the threat of shoulder surfing is not limited to credentials: Surveys revealed that 72% of office commuters in UK shoulder surf the person sitting next to them [6], and globally, 69% of commuters fear that their mobile conversations might be at risk in public settings [13]. Moreover, people have strong concerns about sharing private data on their mobile phones [7]. Suggested solutions to visual privacy include narrowing the viewing angle of the display [10], blinding [15, 18], masking [18] and distorting [3] sensitive information as well as scanning the user's surroundings for potential shoulder surfers and displaying an alert signal [1, 18].

In contrast to prior work, the concept we present in this abstract takes advantage of the “human aspect” in usable security: We hypothesize that replacing the standard print font commonly used on mobile devices with the user's handwriting can protect sensitive textual data from shoulder surfers since users will be able to read their handwriting faster and easier than strangers.

We conducted a first user study with twelve participants as proof of concept for future work and investigated whether there is a difference in reading time between the participants' own handwriting and the handwriting of the other participants. As a baseline we also examined the reading time for a sans-serif print font as commonly used on mobile devices. The results support our approach: Reading time was significantly longer for unfamiliar handwriting compared to the participants' own handwriting ($p < 0.001$), and participants tended to make more errors. Consequently, replacing the print font with the user's handwriting might indeed offer an effective protection against shoulder surfers while maintaining an acceptable level of legibility for the user.

Threat Model

We assume that a user is reading sensitive text messages like chat conversations or emails on a mobile device in a public setting, such as a crowded cafe or train. A stranger standing or sitting next to the user has perfect sight on the device's screen, that is without any reflections or occlusions. Shoulder surfing occurs in a non-malicious way and within limited time, but it reveals sensitive information that the user does not want to share.

Concept

Our idea is to protect sensitive textual content like chat messages from shoulder surfers by replacing the standard print font used on mobile devices with the user's handwriting (see Figure 1). Research on handwriting and fonts has shown that handwriting is less legible than print and that it takes more time to read [11]. This property alone would not improve the security of the user's data, but handwriting has another characteristic which makes it more interesting than simply replacing the print font with a computer cursive font: its individuality [14]. We thus assume that a user is familiar with his or her own handwriting while a stranger is not. Familiarity with a font plays an important role when it comes to reading speed [2]: Familiar fonts are read faster than unfamiliar ones. Consequently, we hypothesize that replacing the print font with the user's handwriting will affect legibility *unevenly*: While the user will still be able to read his or her handwriting with acceptable speed and

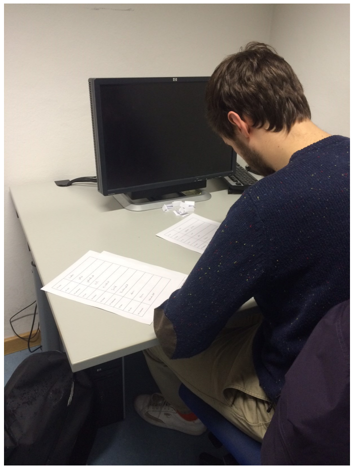


Figure 2: The setting of the first session of our user study.

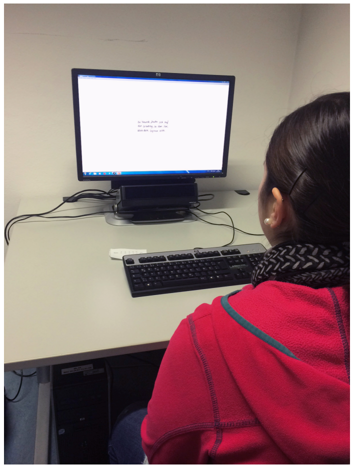


Figure 3: The setting of the second session of our user study.

error rate, deciphering the unfamiliar font will be difficult for strangers, will slow down the reading speed and lead to more reading errors, and thus prevent the understanding of the content's meaning.

User Study

We conducted a proof-of-concept user study in order to examine whether there are differences in reading time between the participants' own handwriting and the handwriting of the other participants in the first place. We used a within-participant repeated-measures design. The dependent variable was *reading time* in milliseconds, the independent variable was *font* with the three levels *own*, *other* and *print*.

Participants

Twelve participants, aged between 21 and 32, were recruited via social networks and personal invitation. In order to avoid bias caused by possible differences between male and female handwriting, gender was counterbalanced. All participants were native German speakers and right-handed. They were either students or staff members, the education level ranged from A Level to PhD.

Method

The study consisted of two sessions a week apart. Both sessions took place in a separate room at our premises. We used the German version of Radner's reading charts [12], a collection of 40 sentences comparable in structure, word and syllable count and difficulty (see Figure 5 a) on page five). Participants were compensated with a 15 Euro shopping voucher.

First session

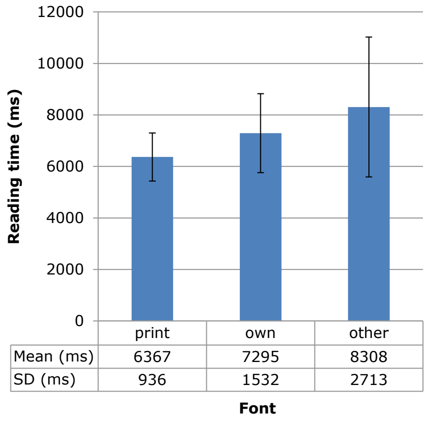
In the first session, we collected handwriting samples of each participant on paper (see Figure 2). Radner's sentences were split into single words in order to avoid memory effects in the second session, and given to the participants as a list in random order. To prevent participants from writing particularly beautifully under the study conditions, they were told to write as if they were jotting down notes.

Second session

For the second session, the words were scanned, cut, and automatically recombined into the original sentences by an algorithm joining the images of the words. For each participant we then built three disjoint sentence sets with eleven sentences each: *own* handwriting, *other* handwriting and *print*. Each participant thus had to read 33 sentences in total. The *print* set consisted of eleven sentences in a sans-serif font, the *other* set included one sentence of each of the other eleven participants. All sentences were scaled to a font size of sixteen points.

During the session, the 33 sentences were presented successively in the center of a white screen. We used a desktop computer setting for the reading task (see Figure 3) since the aim of this study was to test our basic assumption in the first place and to investigate any differences in legibility between the three sentence sets under idealized (e.g. stable) conditions. This setting also allowed for the installation of an eyetracker as a means to check whether the participants followed the instructions and read the whole sentences. Because of the eyetracker, the viewing distance to the display was approximately 70 centimeters (approx. 28 inches). A vision test ensured that all participants were able to read the displayed text from this distance. However, the participants were allowed to find a comfortable reading position within the eyetracker's range, so that the test took place under optimal viewing conditions.

Participants were first presented with a blank white screen and were asked to press space to display a sentence and to start the timer. They should then start reading aloud im-



	print	own	other
Mean (ms)	6367	7295	8308
SD (ms)	936	1532	2713

Figure 4: Means and standard deviations of the reading times per font (calculated using non-transformed data). All pairwise comparisons were significant.

mediately, and press space again as soon as they had read the whole sentence. This stopped the timer and displayed the blank white screen again.

While a participant was reading, the instructor kept a written live record of any reading errors and divided each sentence the participant read into three coarse categories: *“sentence immediately correct”* if the participant read without errors, *“sentence correct after correction”* if the participant stumbled on or misread words, but corrected the errors afterwards, and *“sentence not correct”* if the participant misread words without correcting them, so that the meaning of the sentence was distorted.

Results

In the following sections, we present the results of our user study. A check for normality using Q-Q plots and the K-S test revealed non-normality in the data. Normality was corrected by applying a reciprocal transformation to the data. Unless specifically mentioned otherwise, the following analyses are based on the transformed data.

Reading performance

A repeated-measures ANOVA was conducted to analyze the influence of the font type on reading time. We identified two within-subjects factors: *font* with three levels (*own*, *other*, *print*) and *repetition* with eleven levels (per font type). Mauchly’s test indicated that the assumption of sphericity had been violated. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ϵ = 0.43 for *repetition*, ϵ = 0.56 for *font*).

Repetition did not have a significant effect on reading time, indicating that there was no training effect over the course of the session, *F*(4.24, 46.64) = 1.84, *p* = 0.13. The type of font however had a significant effect on reading time, *F*(1.1, 12.23) = 25.72, *p* < 0.001. Post-hoc tests using the Bonferroni correction revealed that the difference in reading time was significant for all pairwise comparisons (*p* < 0.001 for *other* – *own*, *p* < 0.001 for *other* – *print*, *p* = 0.029 for *own* – *print*). Reading time was shortest for *print* and longest for *other* (see Figure 4). Thus, participants were slower when reading their own handwriting than when reading the printed sentences, but could also read their handwriting faster than that of the other participants. For two participants however the latter was not true, as a comparison of the median *m* of the measured times showed (*m*(*own*) = 6604 ms versus *m*(*other*) = 6539 ms and *m*(*own*) = 7914 ms versus *m*(*other*) = 6682 ms, calculated using non-transformed data).

Reading errors

For the analysis of the reading errors the three categories were labeled as follows: *“sentence immediately correct”* = e_1,

	e_1	e_2	e_3
print	9.5	1.5	0
own	9	1.5	0
other	7.5	2.5	0.5

Table 1: Medians of the reading errors for each font type and error category.

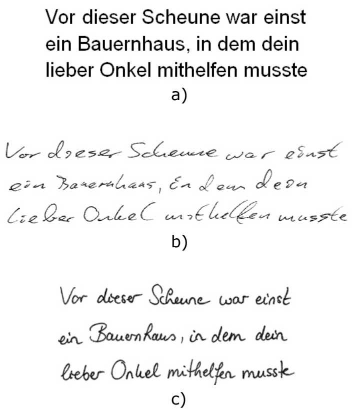

Vor dieser Scheune war einst
ein Bauernhaus, in dem dein
lieber Onkel mithelfen musste
a)
b)
Vor dieser Scheune war einst
ein Bauernhaus, in dem dein
lieber Onkel mithelfen musste
c)

Figure 5: a) An example Radner sentence: “In front of this barn, there was once a farmhouse in which your dear uncle had to help”. b) A handwriting the participants read particularly slow. c) A handwriting the participants read particularly fast.

“sentence correct after correction” = e_2, *“sentence not correct”* = e_3. Grouping according to font type yielded nine different categories: {e_1, *own*}, {e_2, *own*}, {e_3, *own*}, {e_1, *other*} etc.

Using non-transformed data, we then calculated the median *m* for the recorded errors for each of these nine categories. As shown in Table 1, the font type did not lead to errors distorting the meaning of a sentence in most cases (e_3), but participants made more errors when reading the handwriting of the other participants than when reading their own handwriting or the printed sentences. Interestingly, there was no difference between *print* and *own* for e_2 and e_3.

Differences between the participants’ handwriting
Figure 6 shows the mean *M* of the reading times of all participants for sentences written by each participant (x-axis), calculated using non-transformed data. The results imply that there were differences between the participants’ handwriting concerning legibility. For example, when reading the handwriting of participant five (see Figure 5 b)), the other participants were particularly slow, *M* = 10856.18 ms, *SD* = 3064.11 ms, while they were particularly fast when reading the handwriting of participant four (see Figure 5 c)), *M* = 6274 ms, *SD* = 537.93 ms. This suggests that the effect of the approach is dependent on the overall legibility of a handwriting: A scrawly handwriting is better suited to protect sensitive data than a clear and clean one.

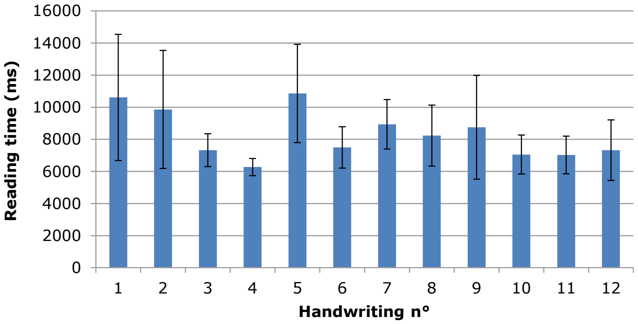


Figure 6: Means and standard deviations of the reading times of all other eleven participants for a particular handwriting (calculated using non-transformed data).

Discussion

In our user study, there was a significant difference between the reading times of the participants’ own handwriting and the handwriting of other participants. Moreover, participants tended to make more errors when reading the unfamiliar handwriting than when reading their own handwriting. Thus, the results suggest that replacing the standard print font used on mobile devices with the user’s handwriting could offer an effective protection of sensitive textual data against casual and non-malicious shoulder surfers. Even though our assumption was tested using a desktop computer setting and the effect was not present for every participant, we argue that the study outcome indicates feasibility for real-world use and provides ground for further study.

Application to a real-world mobile context
In the user study our concept was tested under optimal conditions: The font size was big enough for the participants to read comfortably, participants could choose a comfortable reading distance and they had a direct, unhindered view on the large desktop screen. In contrast, in real-world mobile settings the screen of the device is much smaller and shoulder surfers seldom have perfect sight on the display. Also, in most cases the user and the shoulder surfer will read from different distances: While a user can hold his or her device so that he or she can read comfortably, the reading distance of a shoulder surfer will presumably always be larger. For these reasons, we assume that the

effect observed for our approach is stronger in real-world settings. The differences in reading time between the user and a stranger could be particularly useful for notifications and short text messages. Such content is usually displayed briefly and does not take long to read, so that the user could take full advantage of his or her superior reading speed.

Trade-off between usability and security

Our results suggest that reading time is dependent on the overall legibility of a user's handwriting, pointing to a trade-off between the usability and the security of our approach: A user could either increase the protection level by writing deliberately illegibly or enhance usability by using a clearer handwriting when setting his or her font. Finding the sweet spot between these two factors requires further study.

Technical realization

In contrast to the technical approach utilized in our user study where all sentences were predefined, handwritten-looking sentences would have to be generated dynamically from a user's handwriting samples in a real-world application. A technical realization for mobile devices could build on the prototype developed by Buschek et al. [4] that uses handwriting samples for chat conversations with handwritten-looking fonts.

Limitations

Although the study was thoroughly and carefully designed, we are aware of some limitations which we would like to discuss. One limitation of our study is the small sample size. Moreover, the study was conducted using a desktop computer setting and not a mobile one, and handwriting samples were collected on paper. Thus, this first user study may not necessarily be ecologically valid regarding the envisaged threat model. However, it yields valuable insights on which we can build in future work: It confirmed our basic assumption under controlled, ideal viewing conditions, and although the effect was not observed for all participants, we argue that our results are still promising for real-world use where viewing conditions are seldom perfect for shoulder surfers.

Conclusion and Future Work

In this work, we investigated whether we can take advantage of the individuality of handwriting in order to protect sensitive textual data from shoulder surfers. Assuming that a user is familiar with his or her handwriting while a stranger is not, we showed that utilizing a user's handwriting instead of a print font could indeed enhance visual privacy. The results of a proof-of-concept user study revealed significant differences in reading times: Our participants were slower and more likely to make errors when reading the unfamiliar handwriting of the other participants than when reading their own handwriting. This advantage could be particularly valuable for briefly displayed content like notifications. However, the outcome of our study also suggests that our concept implies a trade-off between usability and security, since reading time was dependent on the overall legibility of a handwriting. Finding the sweet spot between these two factors remains to be investigated. In future work, we plan to address this issue and conduct an in-depth investigation of our concept in a realistic mobile setting on a larger scale.

Acknowledgments

We greatly thank Prof. Dr. Wolfgang Radner for providing us with his reading charts for this research. We also thank Christian Mai for his valuable input.

References

[1] Mohammed Eunus Ali, Tanzima Hashem, Anika Anwar, Lars Kulik, Ishrat Ahmed, and Egemen Tanin. 2014. Protecting mobile users from visual privacy attacks. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (2014), 1–4. DOI: http://dx.doi.org/10.1145/2638728.2638788

[2] Sofie Beier and Kevin Larson. 2013. How does typeface familiarity affect reading performance and reader preference? *Information Design Journal* 20, 1 (2013), 16–31. DOI: http://dx.doi.org/10.1075/idj.20.1.02bei

[3] Michael Boyle, Carman Neustaedter, and Saul Greenberg. 2009. *Media Space 20+ Years of Mediated Life.* Springer, Chapter Privacy Factors in Video-based Media Spaces, 97–122. http://www.springer.com/computer/hci/book/978-1-84882-482-9

[4] Daniel Buschek, Alexander De Luca, and Florian Alt. 2015. There is more to Typing than Speed: Expressive Mobile Touch Keyboards via Dynamic Font Personalisation. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (2015), 125–130. DOI: http://dx.doi.org/10.1145/2785830.2785844

[5] Marian Harbach, Emanuel von Zezschwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. *Proceedings of the Tenth Symposium On Usable Privacy and Security* (2014), 213–230. https://www.usenix.org/conference/soups2014/proceedings/presentation/harbach

[6] Iron Mountain. 2013. Protecting sensitive company information from the commuter snoopers. (2013). http://www.ironmountain.co.uk/Company/Company-News/News-Categories/Press-Releases/2013/October/8.aspx

[7] Amy K. Karlson, A.J. Bernheim Brush, and Stuart Schechter. 2009. Can I borrow your phone? *Proceedings of the 27th international conference on Human factors in computing systems* (2009), 1647. DOI: http://dx.doi.org/10.1145/1518701.1518953

[8] Federico Maggi, Alberto Volpatto, Simone Gasparini, Giacomo Boracchi, and Stefano Zanero. 2011. Poster: Fast, automatic iPhone shoulder surfing. *Proceedings of the 18th ACM conference on Computer and communications security* (2011), 805. DOI: http://dx.doi.org/10.1145/2046707.2093498

[9] Joseph Maguire and Karen Renaud. 2012. You only live twice or the years we wasted caring about shoulder-surfing. *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers* (2012), 404–409.

[10] George Probst. 2000. *Analysis of the Effects of Privacy Filter Use on Horizontal Deviations in Posture of VDT Operators*. Ph.D. Dissertation. Virginia Polytechnic Institute and State University.

[11] Leslie Quant. 1946. Factors Affecting the Legibility of Handwriting. *The Journal of Experimental Education* 14, 4 (1946), 297–316.

[12] Wolfgang Radner, Wilfried Obermayer, Sibylla Richter-Mueksch, Ulrike Willinger, Michaela Velikay-Parel, and Brigitte Eisenwort. 2002. The validity and reliability of short German sentences for measuring reading speed. *Graefe's Archive for Clinical and Experimental Ophthalmology* 240, 6 (2002), 461–467. DOI: http://dx.doi.org/10.1007/s00417-002-0443-5

[13] Regus. 2014. Business people globally describe their ideal work environment. (2014). http://press.regus.com/united-kingdom/download/60669/gbs11theidealworkplaceenvironmentreport_final.pdf

[14] Sargur N. Srihari, Sung-Hyuk Cha, Hina Arora, and Sangjik Lee. 2002. Individuality of handwriting: a validation study. *Journal of forensic sciences* 47, 4 (2002), 856–872. DOI:http://dx.doi.org/10.1109/ICDAR.2001.953764

[15] Peter Tarasewich, Jun Gong, and Richard Conlan. 2006. Protecting private data in public. *CHI '06 extended abstracts on Human factors in computing systems* (2006), 1409. DOI:http://dx.doi.org/10.1145/1125451.1125711

[16] Susan Wiedenbeck, Jim Waters, Leonardo Sobrado, and Jean-Camille Birget. 2006. Design and evaluation of a shoulder surfing resistant graphical password scheme. *Proceedings of the working conference on Advanced visual interfaces* (2006), 177–184. DOI: http://dx.doi.org/10.1145/1125451.1125711

[17] Emanuel von Zezschwitz, Alexander De Luca, Bruno Brunkow, and Heinrich Hußmann. 2015. SwiPIN - Fast and Secure PIN-Entry on Smartphones. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), 1403–1406. DOI: http://dx.doi.org/10.1145/2702123.2702212

[18] Huiyuan Zhou, Vinicius Ferreira, Sao Carlos, Thamara Alves, Kirstie Hawkey, and Derek Reilly. 2015. Somebody Is Peeking! A Proximity and Privacy Aware Tablet Interface. *Extended Abstracts on Human Factors in Computing Systems* (2015), 1971–1976.