Poster: Perceptions of Handling Sensitive Data in Cloud Office Applications

Dominik Wermke, Christian Stransky, Nicolas Huaman, Niklas Busch,
Alexander Krause, Yasemin Acar, and Sascha Fahl
Leibniz University Hannover
Email: {lastname}@sec.uni-hannover.de

***Abstract*—Cloud-based office applications such as Google Docs or Microsoft Office 365 became widely used in recent years, often replacing offline office solutions. The switch to online-hosted office introduces additional security and privacy risks for handling sensitive data and storing documents. Unsuspecting users may not know how, where, or by whom their documents are accessed and stored. Often it is unclear how comfortable end-users are with handling sensitive data in this context.**

In this poster, we present work in progress investigating the interaction of end-users with cloud office software. For this, we conduct surveys with 200 cloud office users from Germany and the U.S. to investigate their perceptions regarding handling sensitive data in cloud office suites. While still preliminary work, we find that some users' mental models are incomplete and their understanding of cloud office security and privacy is limited, often caused by a lack of transparency of the services' operations.

I. Motivation

In recent years, most major office software providers have moved on to provide some sort of cloud platform. Popular cloud office applications like Microsoft Office 365, Google Drive, and projects like LibreOffice Online allow for collaboration between multiple editors, automatic real-time storage on cloud or internal network servers, and easy access through the browser without requiring the installation of software.

With the shift from offline to cloud, many cloud office providers also moved from a pay-once model to a subscription-based model with a trial period or even a completely free payment model. This shift accompanied a questionable change in business model drive for these companies: the processing and storing of documents in the cloud provides the possibility of large-scale privacy intrusion by the providers for both end-users and businesses that utilize the cloud.

Due to the recent corona virus pandemic this consideration becomes increasingly more relevant, as the sudden emergence of the virus forces millions into self-isolation.

With the enforcement of social distancing due to the corona virus epidemic, many employees are forced to work from home with limited access to their company infrastructure and colleagues. As a result, they often have to rely on cloud-based services to continue work with their colleagues in a home office environment. The sudden, unprepared switch to a home office environment may require the handling of sensitive documents such as contracts and signatures in cloud office software. In addition, the home office environment may require the handling of sensitive documents such as contracts and signatures in cloud office software.

II. Approach

In this work in progress, we investigate privacy and security misconceptions by users of cloud office applications. To investigate the interaction of users with cloud office software, we conducted two online surveys with crowd workers from Amazon's Mechanical Turk ($n = 105$) and ClickWorker ($n = 95$).

A. Survey Structure

Both the German-speaking participants from ClickWorker and the English-speaking participants from Mechanical Turk were administered an almost identical survey, with the German survey being a direct translation from the English one by multiple native German speakers. Differences included slight changes due to localization (e.g., localized names for government agencies) and changes to concepts that do not exist or have a different privacy implication (e.g., social security number). The questionnaire development was guided by our established research questions. We performed 5 in-depth, free-form interviews with both experts and non-experts following the principle of cognitive interviews [1]. In addition, we refined the surveys in multiple pilots with participants on Mechanical Turk ($n = 9$) and ClickWorker ($n = 20$) until a satisfactory convergence was reached.

We did not mention security or privacy in the initial recruitment ad to avoid certain recruitment biases. We generally required participants to be age 18 or older and to have used cloud office software before. For Amazon's Mechanical Turk, we additionally required participants to live in the U.S. To ensure sufficient data quality, we also required them to have completed a minimum of 1,000 hits and to have a task approval rate of at least 95% [2]. For ClickWorker, we additionally required participants to speak German and to live within Germany, Austria, or Switzerland.

Our institutions did not require a formal IRB process for the studies conducted in this work. Nonetheless, we modeled our research plan and study procedures after an IRB approved study, adhered to the strict German and U.S. data and privacy protection laws and the General Data Protection Regulation in the E.U., and structured our study following the ethical prin-

cipals of the Menlo report for research involving information and communications technologies [3].

A total of 229 people responded to our surveys. Of those, 22 did not finish and 7 were excluded due to low-quality answers or due to failing at least one of our quality checks, resulting in 200 final participants whose responses we consider.

B. Scenarios

We presented our participants with three different types of sensitive data potentially handled in cloud office applications: children's data including names and grades, health data including names and diagnosis, and financial data including names and SSNs, either in a more personal or more generalized condition. Participants of the study were equally distributed between both conditions and the order of scenarios was randomized for each participant.

Scenario 1: Children's Data. The first scenario described the use of a cloud office application in an educational setting. We asked our participants to assess how much they felt at ease with using cloud office applications for handling data of children in schools, e.g., for storing grades or writing tasks.

Scenario 2: Health Data. The second scenario had a focus on health information. A general practitioner used a cloud office application to handle sensitive patient information including a patient's name, age, weight, diagnosis, and treatment plan. Again, we asked our participants to rate their level of comfort with the scenario.

Scenario 3: Financial Data. In the third scenario we illustrated a use case involving financial data. A financial advisor used a cloud office application to process client data. The processed documents include private information such as the client's name, social security number, and detailed financial information.

Conditions. Participants were equally (and randomly) distributed across two conditions: "General" scenarios with a more generalized phrasing and "Personal" scenarios with more personalized phrasing (e.g., "a child" vs. "your child").

C. Findings

We report the general demographics of both surveys in Table I. Somewhat unsurprisingly, participants prefer to store their documents on the platform they edit them with (e.g., locally for offline office). Participants of both the U.S. and German survey agree on the top reasons why they (would) use cloud office applications over local office applications: easy remote access of documents (76.2%, 70.5%), ease of collaboration (58.1%, 59.0%), and free or cheap access (52.4%, 43.2%).

In the three scenarios, both the "Health data" scenario and the "Financial data" scenario are rated as less comfortable by our participants than the "Child data" baseline. (cf. Figure 1) Overall, our participants are uncomfortable the most with the scenario of processing financial documents in the cloud. Differences between a more general scenario and a more personalized scenario did not measurably affect our participants' comfort level nor did their survey language.

TABLE I: Demographics for all valid participants from the U.S. survey (Amazon's Mechanical Turk), German survey (ClickWorker), and combined.

	U.S.	German	Combined
Participants			
Started	127	102	229
Finished	110	97	207
Valid ($n =$)	105	95	200
Office Usage*			
Google Drive	97.1%	80.0%	89.0%
Microsoft Office (Offline)	86.7%	87.4%	87.0%
Microsoft Office 365 (Cloud)	70.5%	64.2%	67.5%
LibreOffice Offline	18.1%	25.3%	21.5%
Apple's iWork Web (Offline)	9.5%	20.0%	14.5%
Apple's iWork Web (Cloud)	6.7%	17.9%	12.0%
LibreOffice Online	4.8%	9.5%	7.0%
Other	3.8%	5.3%	3.0%
OnlyOffice	1.0%	1.1%	2.5%

* Multiple answers allowed, may not sum to 100%

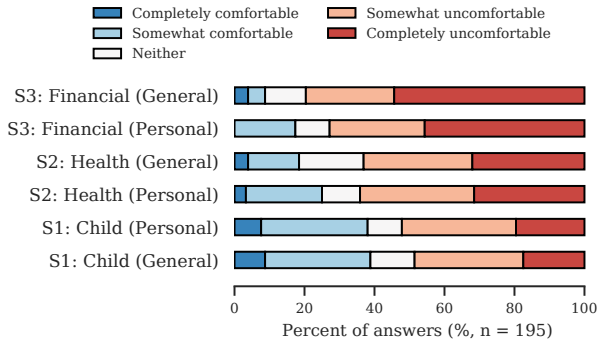

Fig. 1: Participants' comfort with three different data scenarios (Financial, Health, and Children) and two different conditions (General and Personal perspective).

III. Outlook

While still a work in progress, we think our preliminary results will allow for novel insights into the perceptions of cloud office users regarding the handling of sensitive data in cloud office applications. General misconceptions, ambiguous access rights, and the unclear responsibilities of cloud providers seem to provide additional challenges for the end-user adoption of cloud office suites. Especially as the current state of cloud office suites might leave much to be desired in the eyes of end users.

References

[1] S. Presser, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, J. M. Rothgeb, and E. Singer, "Methods for Testing and Evaluating Survey Questions," *Public Opinion Quarterly*, vol. 68, no. 1, pp. 109–130, 03 2004. [Online]. Available: https://doi.org/10.1093/poq/nfh008

[2] E. Peer, J. Vosgerau, and A. Acquisti, "Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk," *Behavior research methods*, vol. 46, 12 2013.

[3] D. Dittrich and E. Kenneally, "The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research," U.S. Department of Homeland Security, Tech. Rep., Aug 2012.