Privacy-preserving Comparison of Cloud Exposure Induced by Mobile Apps

Martin Henze*, Ritsuma Inaba§, Ina Berenice Fink*, Jan Henrik Ziegeldorf*
*Communication and Distributed Systems, RWTH Aachen University, Germany
§College of Engineering, University of Michigan, USA
{henze,fink,ziegeldorf}@comsys.rwth-aachen.de,rinaba@umich.edu

ABSTRACT

The increasing utilization of cloud services by mobile apps on smartphones leads to serious privacy concerns. While users can quantify the cloud usage of their apps, they often cannot relate to involved privacy risks. In this paper, we apply comparison-based privacy, a behavioral nudge, to the cloud usage of mobile apps. This enables users to compare their personal app-induced cloud exposure to that of their peers to discover potential privacy risks from deviation from normal usage behavior. Since cloud usage statistics are sensitive, we protect them with k-anonymity and differential privacy.

CCS CONCEPTS

• **Security and privacy → Database and storage security**; *Human and societal aspects of security and privacy*; • **Human-centered computing → Ubiquitous and mobile computing**;

KEYWORDS

Privacy, Smartphones, Cloud Computing, Security, Anonymity

ACM Reference Format:
Martin Henze, Ritsuma Inaba, Ina Berenice Fink, Jan Henrik Ziegeldorf. 2017. Privacy-preserving Comparison of Cloud Exposure Induced by Mobile Apps. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous 2017).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3144457.3144511

1 INTRODUCTION & MOTIVATION

Mobile apps on smartphones increasingly utilize cloud services [4]. The enormous benefits of this come at the price of serious privacy risks [3]. To uncover cloud exposure caused by mobile apps, CloudAnalyzer [4] dissects network traffic on smartphones to detect cloud usage. However, especially for less technically proficient users, it is extremely difficult to relate to the identified potential privacy risks.

Comparison-based privacy was introduced in the similar context of over-sharing in social media [9] to enable users to compare themselves along different privacy-relevant metrics to their peer groups. This approach obviates the need for fixed privacy norms or ground truth as a basis for building nudges—this is what is also needed in the context of cloud usage. While applying comparison-based privacy to nudge users on mobile apps' cloud usage is extremely promising, it also introduces privacy concerns: (i) the operator of the comparison system could learn the peer groups of users, (ii) the operator could try to infer the identity of users, (iii) the operator could link multiple contributions of users, and (iv) small comparison groups could leak users' contributions or installed apps. Hence, comparison-based privacy requires additional security measures.

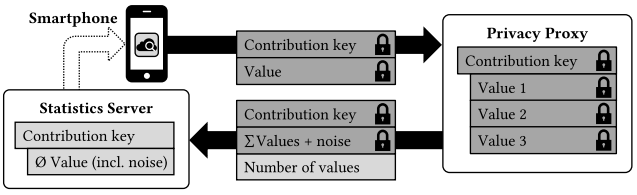

Figure 1: Anonymous comparison of mobile cloud usage

In this paper, we study the feasibility and applicability of *securely* realizing comparison-based privacy on smartphones to nudge users on the cloud usage of their mobile apps. We introduce a privacy proxy that hides users' identities and employs k-anonymity [7] and differential privacy [2] to aggregate and to further protect user contributions from disclosure. Our preliminary feasibility study with 29 volunteers over the course of 19 days shows that comparison-based privacy is a promising approach towards supporting users in exercising their right for privacy when using mobile apps.

2 COMPARISON OF MOBILE CLOUD USAGE

Our proposal for anonymously comparing the cloud usage of mobile apps is shown in Figure 1. As detailed in the following, the *smartphone* collects statistics on cloud usage and periodically sends these statistics in encrypted form to the *privacy proxy*. The privacy proxy—without being able to decrypt the statistics—aggregates statistics of different users and adds random noise before releasing the aggregate to the *statistics server*. The statistics server is able to decrypt the aggregated statistics and provides them back to the smartphone to enable comparison. This process is secure as long as privacy proxy and statistics server do not collude, which can, e.g., be achieved when they are operated by different parties.

Smartphone. We use CloudAnalyzer [4] to detect cloud usage of mobile apps on Android using IP adresses, DNS names, and TLS information in network traffic. Based on the information provided by CloudAnalyzer, the smartphone calculates the *contribution value* for each day and app, i.e., the fraction of traffic that has been sent to cloud services, and encrypts this with the public key of the statistics server using an additive homomorphic cryptosystem. Furthermore, it creates a *contribution key* consisting of the app's name, date, and the peer group. The smartphone then encrypts the contribution key with the statistics server's public key using a deterministic cryptosystem and sends encrypted key and value to the privacy proxy.

MobiQuitous 2017, November 7–10, 2017, Melbourne, VIC, Australia

ACM ISBN 978-1-4503-5368-7/17/11.
https://doi.org/10.1145/3144457.3144511

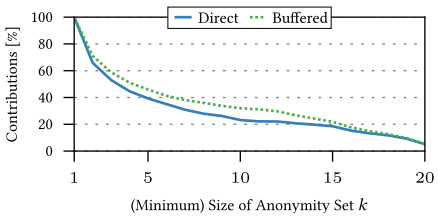

Figure 2: Impact of k-anonymity

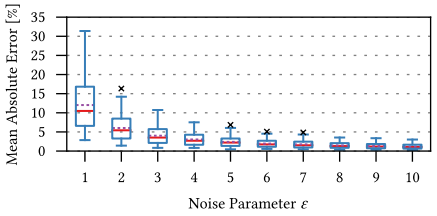

Figure 3: Impact of ε-differential privacy

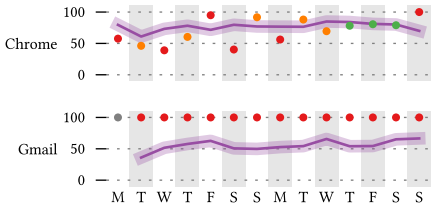

Figure 4: Example of comparison result

Privacy Proxy. While the privacy proxy cannot decrypt received keys and values, it can derive which values belong to the same key (because of the deterministic cryptosystem). To achieve k-anonymity [7], the proxy waits until it received at least k values for the same key. It then adds up the k values (using the additive homomorphic cryptosystem) and adds Laplacian noise to achieve ε-differential privacy [2]. Finally, it releases the still encrypted noisy sum, the number of values, and the key to the statistics server.

Statistics Server. The statistics server decrypts the received key and the noisy sum. It then calculates the noisy mean value by dividing the noisy sum by the number of values. Finally, it stores the key and the mean value in a database.

Users can query the statistics server for the anonymized mean cloud usage for a particular key (app name, date, peer group). The secure combination of privacy proxy and statistics server thereby guarantees the privacy of users and their contributions.

3 FEASIBILITY STUDY & EARLY RESULTS

To assess the feasibility and applicability of our approach, we implemented a prototype for Android. We realized privacy proxy and statistics server with Python and use Paillier [6] as additive homomorphic cryptosystem and a combination of salted SHA-256 hashes with a crypto box construction [1] as deterministic cryptosystem.

We recruited volunteers to record statistics on 29 Android devices during 19 days. Our statistics contain 383 apps and 347 days of mobile device usage. We refer to the CloudAnalyzer paper for a detailed discussion of the study design and ethical considerations [4].

Influence of k-Anonymity. We study the influence of the size of the anonymity set (k) in Figure 2. The choice of k directly influences which contributions can be included in the analysis, as contributions for a specific key (app name, date, peer group) can only be used if at least k users provide their values. Furthermore, the privacy proxy can either *directly* forward contributions as soon as the threshold k is reached or first *buffer* them (e.g., for a day) before releasing data for all keys with $\geq k$ contributions. For the 29 studied devices, Figure 2 shows that 28.9 % of contributions are unique and hence clearly cannot be shared. Buffering contributions (for a day) slightly increases the fraction of usable contributions. For a reasonable choice of $k = 5$ (for our small number of contributors) [8], we can still utilize 39.3 % (direct) resp. 46.0 % (buffered) of the contributions. For larger numbers of users—where more usable contributions are expected—increasing k to 10 is advisable [8]. For our small dataset, we fix $k = 5$ and buffer contributions for one day in the following.

Influence of ε-Differential Privacy. To study the impact of differentially private noise, we replay the data collected by our volunteers 30 times using random seeds to generate Laplacian noise for different privacy parameters ε. Figure 3 shows the distribution of the mean absolute error for each app and day (over 30 runs) for different ε. The challenge here is to add noise such that privacy is protected and the result is still usable. For $\varepsilon = 1$, the mean absolute error on average amounts to 12.0 % (dotted line), which clearly impacts utility. In contrast, $\varepsilon = 5$ with a mean absolute error of on average 2.4 % provides a good trade-off between privacy and utility for our small dataset. We hence use $\varepsilon = 5$ in the following.

Comparison Result. Figure 4 exemplarily shows the comparison of one of our volunteers to their peer group. The violet line represents the anonymized mean cloud usage within the peer group with a 10 % margin. Over a period of 2 weeks, each dot shows the cloud usage of the user on a particular day. Here, colors inform the user how much their cloud usage deviates from the peer group. For our volunteer, we observe that the usage pattern is quite similar to the peer group for the `Chrome` app. However, for `Gmail` (standard email app on Android) the volunteer's cloud usage is significantly higher than in the peer group, identifying potential privacy risks.

4 CONCLUSION & OUTLOOK

Relating to the privacy risks of app-induced cloud exposure significantly challenges less technically proficient users. Our preliminary results indicate that anonymously comparing the cloud usage of mobile apps is indeed a feasible and promising approach to nudge users to exercise their right for privacy. Further more, we believe that our approach is also valuable to uncover cloud exposure in other use cases (e.g., email [5]) and to study other privacy aspects of mobile device usage beyond cloud exposure (e.g., location sharing). In the future, we will test and validate our approach in a larger study.

Acknowledgments. The authors would like to thank the participants of the user study. This work has received funding from the German Federal Ministry of Education and Research (BMBF) under project funding reference no. 16KIS0351 (TRINICS). The responsibility for the content of this publication lies with the authors.

REFERENCES

[1] D. J. Bernstein. 2009. *Cryptography in NaCl.* Technical Report. UIC.
[2] C. Dwork. 2006. Differential Privacy. In *ICALP 2006.*
[3] M. Henze et al. 2016. Towards Transparent Information on Individual Cloud Service Usage. In *IEEE CloudCom 2016.*
[4] M. Henze et al. 2017. CloudAnalyzer: Uncovering the Cloud Usage of Mobile Apps. In *MobiQuitous 2017.*
[5] M. Henze et al. 2017. Veiled in Clouds? Assessing the Prevalence of Cloud Computing in the Email Landscape. In *IEEE/IFIP TMA 2017.*
[6] P. Paillier. 1999. Public-key cryptosystems based on composite degree residuosity classes. In *EUROCRYPT 1999.*
[7] L. Sweeney. 2002. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 5 (2002).
[8] K. Wiesner et al. 2013. Privacy-Preserving Calibration for Participatory Sensing. In *MobiQuitous 2013.*
[9] J. H. Ziegeldorf et al. 2015. Comparison-based Privacy: Nudging Privacy in Social Media (Position Paper). In *DPM 2015.*