Outline: An Extensive and Secure Personal Data Management System Using SGX

Robin Carpentier, Floris Thiant, Iulian Sandu Popa, Nicolas Anciaux, Luc Bouganim

▶ **To cite this version:**

Robin Carpentier, Floris Thiant, Iulian Sandu Popa, Nicolas Anciaux, Luc Bouganim. Outline: An Extensive and Secure Personal Data Management System Using SGX. 1st CNIL International Privacy Research Day, Jun 2022, Paris, France. . hal-03763815

HAL Id: hal-03763815
https://inria.hal.science/hal-03763815v1

Submitted on 29 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Outline: An Extensive and Secure Personal Data Management System Using SGX

Robin Carpentier
Univ. Versailles St-Q.-en-Yvelines
robin.carpentier@uvsq.fr

Floris Thiant
Inria Saclay
floris.thiant@inria.fr

Iulian Sandu Popa
Univ. Versailles St-Q.-en-Yvelines
iulian.sandu-popa@uvsq.fr

Nicolas Anciaux
Inria Saclay
nicolas.anciaux@inria.fr

Luc Bouganim
Inria Saclay
luc.bouganim@inria.fr

This outline was submitted alongside a demonstration paper [3] to be accepted for a presentation at the 2022 CNIL's International Privacy Research Day [4].

1 ABSTRACT

Personal Data Management System (PDMS) solutions are currently flourishing, spurred by new privacy regulations such as GDPR and new legal concepts like data altruism. PDMSs aim to empower individuals by providing appropriate tools to collect and manage their personal data and share computed results with third parties, thus requiring (i) a secure platform protecting the user's privacy and delivering strong guarantees on the outputs of user's data processing, and (ii) an extensible solution that supports all types of data-driven computations.

2 GOAL

Our objective is to give individuals the means to benefit from data portability in a greater way, by safeguarding their control on the retrieved data. Instead of transmitting the retrieved data to a third party for further use, we want to provide solutions that allow the user to securely retrieve the data and perform computations on behalf of a third party, while controlling the result disclosed to that third party. On the one hand, individuals need automatic tools to collect their data from the multiple data sources they use. On the other hand, after exercising their right to data portability, they need to allow third parties to carry out computations without giving access to the raw data. We want to provide an Extensive and Secure PDMS facilitating these aspects for layman users.

3 SCENARIOS

Consider the following two scenarios as examples of computations requiring the disclosure of aggregates to third parties:

Green bonus example. Companies want to reward their employees for having an ecological conduct. Thus, they monthly compute a green bonus (financial incentive) based on the number of commutes by bike. A user-centered workflow is possible: GPS traces are collected from a reliable service (e.g., Google Maps), then processed locally (e.g., on the user's PDMS) by a specific code provided by the employer. The result is then delivered to the employer with a proof of compliance.

Energy example. An energy supplier wishes to offer services precisely calibrated to the energy consumption of its future customers. To establish a tailor-made offer, the supplier must evaluate various statistical computations on the customer's consumption. Using the data collected from their energy supplier (or smart meter) and thanks to their PDMS offering confidentiality guarantees, customers agree to disclose these statistics but not their detailed consumption and thus install a computation function sent by the supplier. The attestations provided by the PDMS allow the supplier to commit to a quote since they get guarantees on the computation performed.

4 METHODOLOGY

To handle these scenarios, we built a PDMS architecture providing: i) confidentiality guarantees for the individual, so that the third parties will not have access to their raw data; ii) guarantees for the third parties that the computation was performed by a specific program they specify, run on the required personal data, so that the user cannot cheat and they can commit to the financial incentive (green bonus) or quote (energy) ; iii) extensibility, so that the computation code can provided by the third party is not constrained and remains generic enough. We leverage Trusted Execution Environments (TEE) combined with sandboxing techniques to ensure the security of any user device likely hosting the PDMS [1]. We also rely on execution strategies that control execution flow and size of intermediate results to mitigate the leakage of large amounts of information potentially sent to the third party, as detailed in previous publication [2]. This demonstration paper describes a PDMS implementation using Intel SGX and provides a set of games to ease the comprehension of the proposal.

5 EXISTING WORKS

Current practical solutions would resort to web services to collect personal data and exploit it, which would raise privacy concerns and a perceived risk of mass surveillance. PDMS solutions (such as CozyCloud or Digi.me) would fail to provide the necessary security and extensiveness properties, as they only rely on (community) semantic analysis of computation code to ensure security. Similarly, secure database outsourcing techniques, which rely on encrypting personal data before sending it to the third-party service to perform calculations on the encrypted data, would not allow any (generic) processing. Similarly, state of the art solutions in automatic information flow analysis (computation code/results analysis) would not be applicable here, as they are based on an assumption of non-interference between any sensitive inputs and the computation result, whereas in our scenarios, the aggregated (statistical) results can obviously depend on any personal input.

REFERENCES

[1] N. Anciaux, P. Bonnet, L. Bouganim, B. Nguyen, P. Pucheral, I. Sandu Popa, and G. Scerri. 2019. Personal Data Management Systems: The security and functionality standpoint. *Inf. Syst.* 80 (2019), 13–35.

[2] R. Carpentier, I. Sandu Popa, and N. Anciaux. 2021. Poster: Reducing Data Leakage on Personal Data Management Systems. In *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE Computer Society, 716–718.

[3] Robin Carpentier, Floris Thiant, Iulian Sandu Popa, Nicolas Anciaux, and Luc Bouganim. 2022. An Extensive and Secure Personal Data Management System Using SGX. In *Proceedings of the 25th International Conference on Extending Database Technology, EDBT 2022*. 2:570–2:573. https://doi.org/10.48786/edbt.2022.53

[4] Robin Carpentier, Floris Thiant, Iulian Sandu Popa, Nicolas Anciaux, and Luc Bouganim. 2022. An Extensive and Secure Personal Data Management System Using SGX. Presentation at the 1st International Privacy Research Day, Commission Nationale de l'Informatique et des Libertés, Paris, France. https://www.cnil.fr/en/privacy-research-day-discover-program-first-cnils-international-conference