

Sedigheh Eslami, Asia Biega, Rishiraj Saha Roy, Gerhard Weikum

Max Planck Institute for Informatics

Motivations

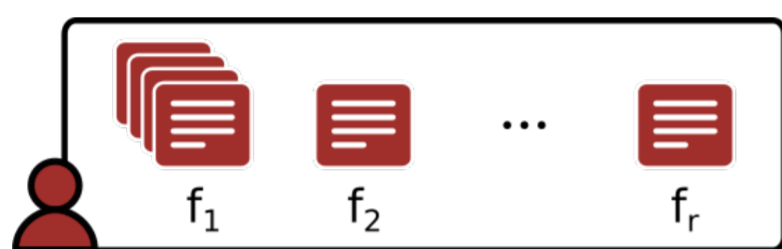
1. Service providers keep removed user profiles for internal data analysis.
2. Hiding removed profiles is not enough and can compromise users privacy e.g in case of hacking or governmental inquiries.

Objective

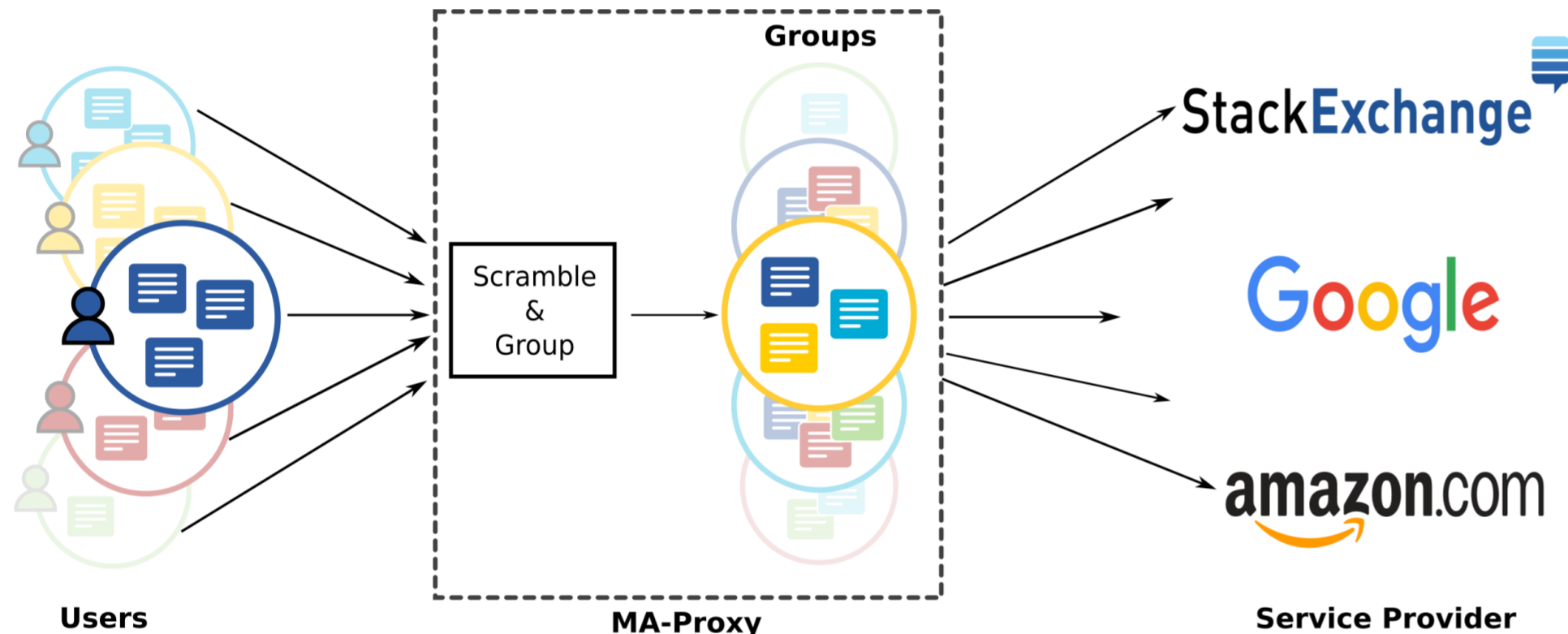
- ▶ Finding a better method for profile removal that preserves users privacy *and* minimizes the service provider utility loss.

Mediator Accounts and Profile Removal

- ▶ User Profile Schema



- ▶ Mediator Accounts Architecture



- ▶ Attack Model



Measurements

- ▶ User Privacy Gain

- ▶ Set of deleting user profiles: $\mathbf{DU} = \{u_1, \dots, u_l\}$,
- ▶ Set of grouped accounts: $\mathbf{G} = \{g_1, \dots, g_m\}$,
- ▶ Groups probability vector per user: $\vec{\phi}_i = \langle \phi_{1i}, \phi_{2i}, \dots, \phi_{mi} \rangle$

$$\text{UPG} = \min_i (\mathbf{H}_i = - \sum_{j=1}^m \phi_{ji} \times \log_2 \phi_{ji}) \quad (1)$$

- ▶ Provider Utility Loss

- ▶ Set of forums $\mathbf{F} = \{f_1, f_2, \dots, f_r\}$,
- ▶ Number of users of type (\cdot) contributing in forums i and j : $u^{(\cdot)}(f_i, f_j)$,

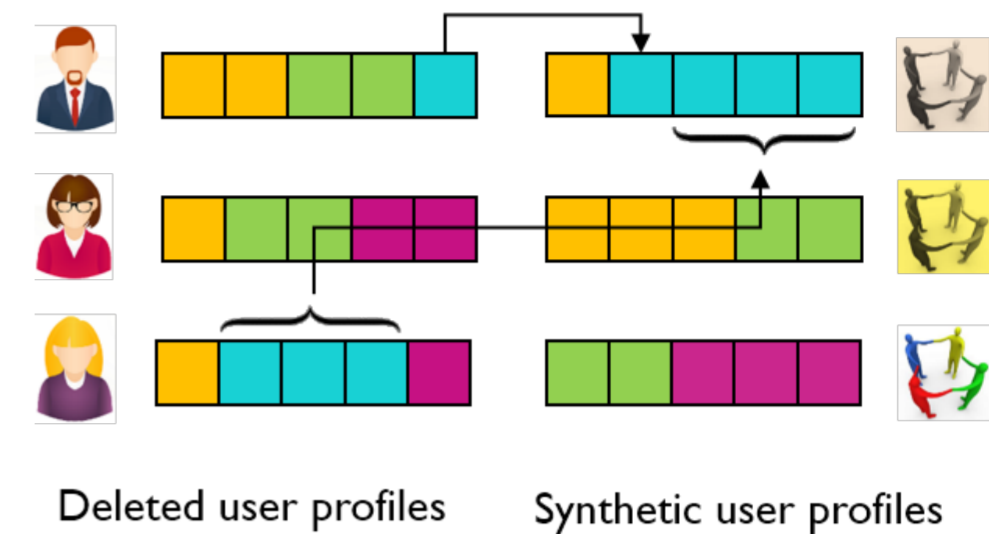
$$\text{PUL} = \frac{\sum \sum_{i,j,i < j} (u^{\text{actUdel}}(f_i, f_j) - u^{\text{actUmed}}(f_i, f_j))^2}{C(r, 2)} \quad (2)$$

Contact Information

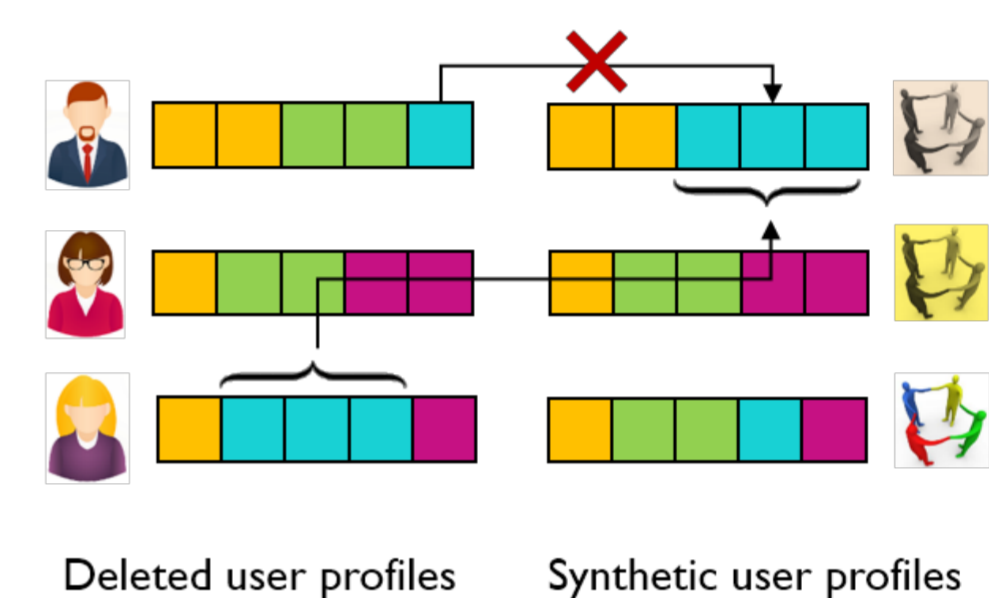
- ▶ sedigheh.eslami@hpi.de
- ▶ {jbiega,rishiraj,weikum}@mpi-infmpg.de

Methods

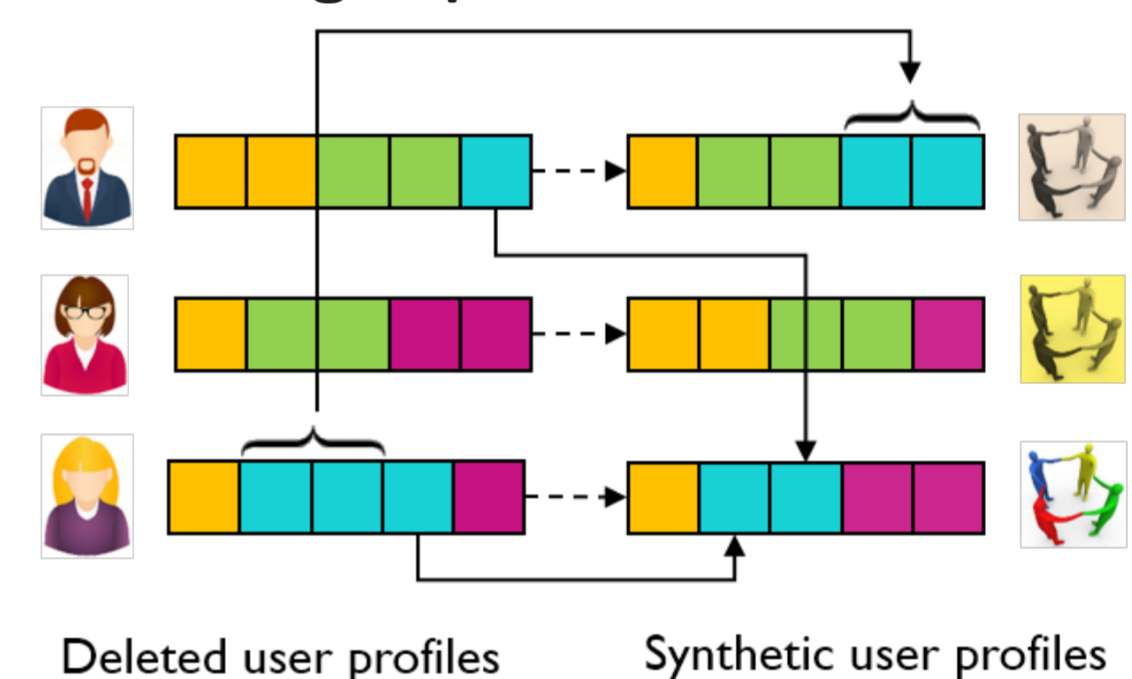
- ▶ Random: *Uniform random* post assignment to grouped accounts
- ▶ Unconstrained: *Split* posts based on forums to make chunks, *merge* chunks from different users.



- ▶ Constrained: Put constraint on *merging* chunks, in order to spread the forum contents.



- ▶ Structured: Preserve structure of the user profiles in terms of forum contribution, in grouped accounts.



Results

- ▶ Experiments on two datasets

Dataset	StackExchange		Healthforums	
Method	Priv-gain	Util-loss	Priv-gain	Util-loss
Random	13.574	3.39×10^6	13.117	6.13×10^5
Unconstrained	6.510	1.12×10^3	13.021	3.59×10^4
Constrained	6.580	9.53×10^2	13.021	2.20×10^4
Structured	13.436	Zero*	13.117	Zero*
Original	0.000	0.000	0.000	0.000

Conclusion

- ▶ Introduction of a novel problem: Privacy of the hidden profiles.
- ▶ Providing an alternative solution to the traditional profile removal:
 - ▶ Reorganizing deleting profiles to synthetic mediator accounts,
 - ▶ Satisfying profiling privacy of users, *while* providing accurate analytics for service providers.

References

- [1] Asia Biega, Rishiraj Saha Roy, and Gerhard Weikum. Privacy through Solidarity: A User-Utility-Preserving Framework to Counter Profiling. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, 2017 (to appear).
- [2] Justin Brickell and Vitaly Shmatikov. The cost of privacy: destruction of data-mining utility in anonymized data publishing. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 70–78. ACM, 2008.