Guaranteed Confidentiality and Efficiency in Crowdsourcing Platforms

https://crowdguard.irisa.fr/

Joris Duguépéroux

PhD student, starting from September 2017, supervised by Allard Tristan and Gross-Amblard David

joris.dugueperoux@irisa.fr

Motivation

Crowdsourcing platforms and gig economy: a new way of working, for good...

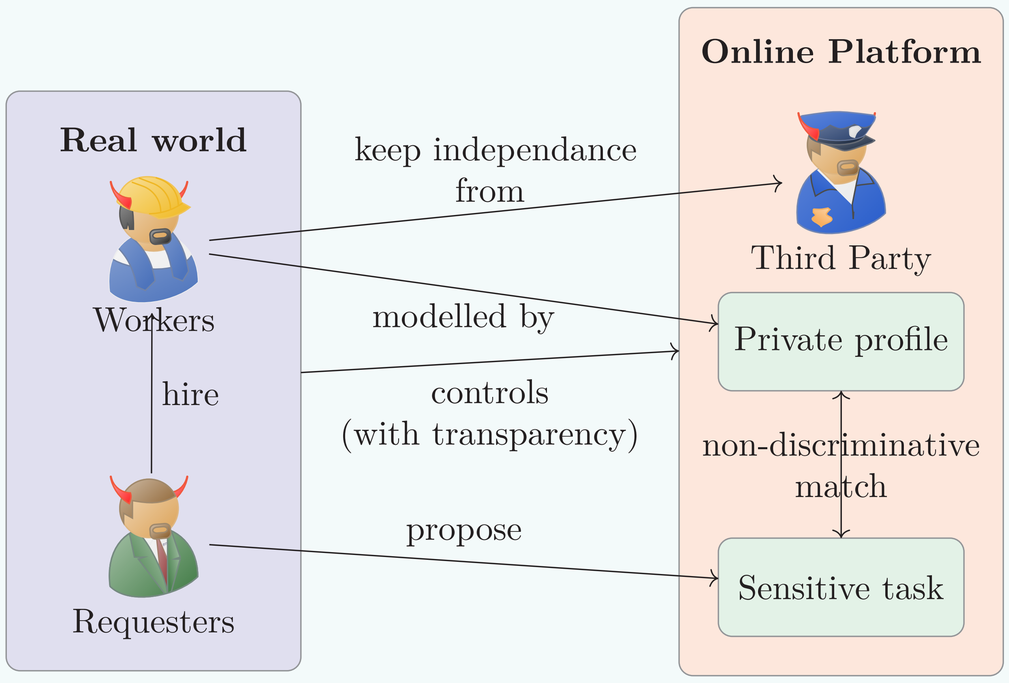
- Wide and easy access to workers and to work
- Potential innovation accelerator (*e.g.,* Kicklox or Tara)
- New possibilities for research (*e.g.,* Foldit)

... and for bad...

- Privacy scandals (*e.g.,* illegitimate accesses to the real-time geolocation traces of riders/drivers (https://tinyurl.com/wp-priv) or de-anonymization[1])
- Denial of workers' independence (see *e.g.,* the complaints of micro-task workers (https://tinyurl.com/wsj-ind))
- Discrimination (*e.g.,* in Uber[2])
- And, yes, major societal/legal issues ...

Long-term objectives of my PhD thesis

- **Individual**
  - protection of workers
  - protection of tasks
  - protection of requesters
- **Collective**
  - Enable independence
  - Avoid discrimination
  - Enforce transparency
  - Enable control
- **Quality and efficiency**
  - Computation time
  - Quality of matching
  - Other quality measures



Privacy techniques in our toolbox

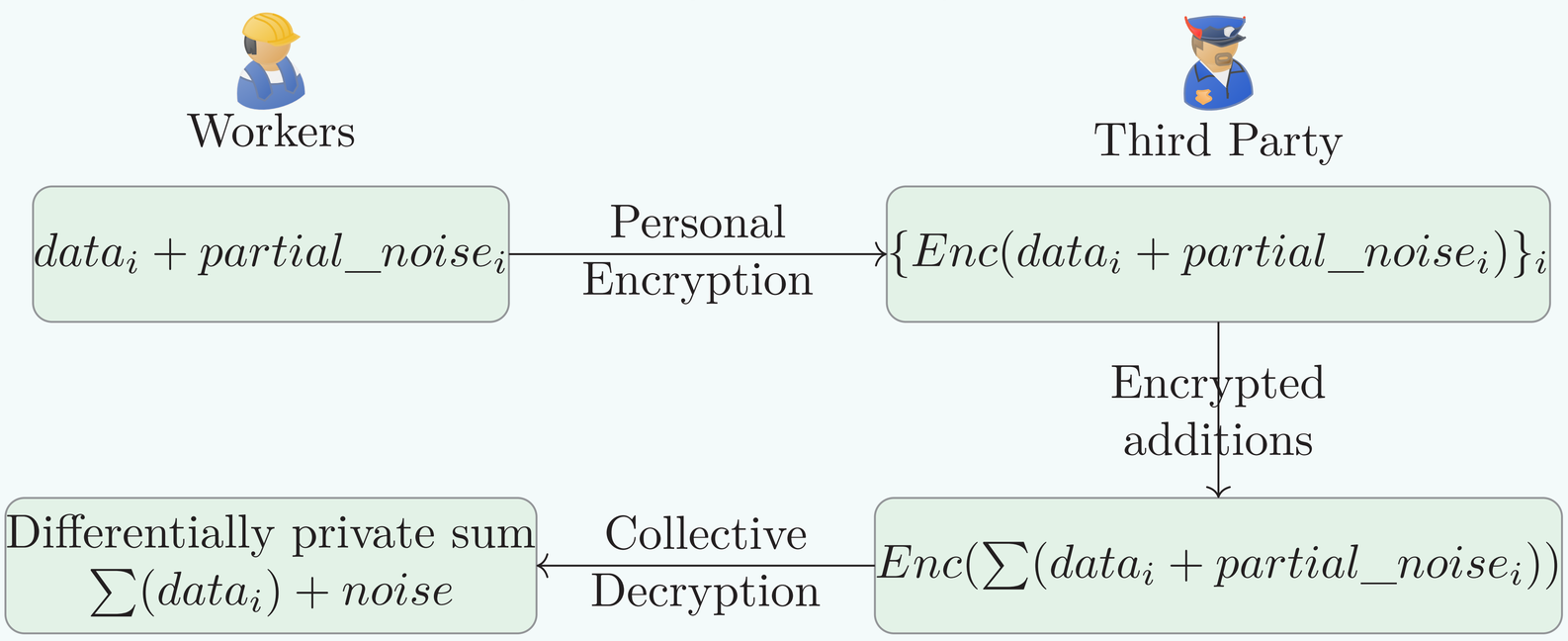
Differential Privacy

- Ensures strong privacy guarantees
- Easy composition with other algorithms
- Gives noisy answers to queries (see application in crowdsourcing [3])

Homomorphic Cryptography

- Strong cryptographic guarantees
- Only reveals the results of operations
- Common decryption (via key-shares)
- Heavy computation time (see application in crowdsourcing [4])

Combining Techniques, Example: Computation of a private sum



Overview of our approach

Our work: Privacy in Assignment

- For workers
- With no trusted third party
- Realistic quality and computation time

Pre-Assignment: Partitioning the space of workers

- With a KD-tree for instance
- Using private sum as building block
- Medians are approximated with noisy histograms

Assignment: Assigning tasks to partitions

- Assigning a task to a set of partitions
- Assignment is public
- Quality measure depends on the data model

Post-Assignment: Enabling contact

- Workers select assigned tasks they like...
- ... and contact them through a private channel (*e.g.,* TOR)

Challenges

Data Model:

- Relevant data model? (*e.g.,* add preferences for workers?)
- Constraints adapted to the issue? (*e.g.,* matching that minimizes distance? threshold for skills?)

Privacy in Assignment:

- Reducing computation time
- Finding ways to evaluate correctly
- Optimizing the differential privacy budget

Extend privacy guarantees:

- Formalize some desired properties (*e.g.,* independence of workers, transparency)
- Choosing an appropriate definition when several co-exist (*e.g.,* discrimination)
- Formalize the attacker model

Experiments: Realistic assumptions on data distribution

References

[1] Matthew Lease, Jessica Hullman, Jeffrey P Bigham, Michael S Bernstein, Juho Kim, Walter Lasecki, Saeideh Bakhshi, Tanushree Mitra, and Robert C Miller. Mechanical turk is not anonymous. 2013.

[2] Yanbo Ge, Christopher R Knittel, Don MacKenzie, and Stephen Zoepf. Racial and gender discrimination in transportation network companies. Technical report, National Bureau of Economic Research, 2016.

[3] Louis Beziaud, Tristan Allard, and David Gross-Amblard. Lightweight privacy-preserving task assignment in skill-aware crowdsourcing. In *To appear in the 28th International Conference on Database and Expert Systems Applications (DEXA '17)*, 2017.

[4] Hiroshi Kajino. *Privacy-Preserving Crowdsourcing*. PhD thesis, 2015.