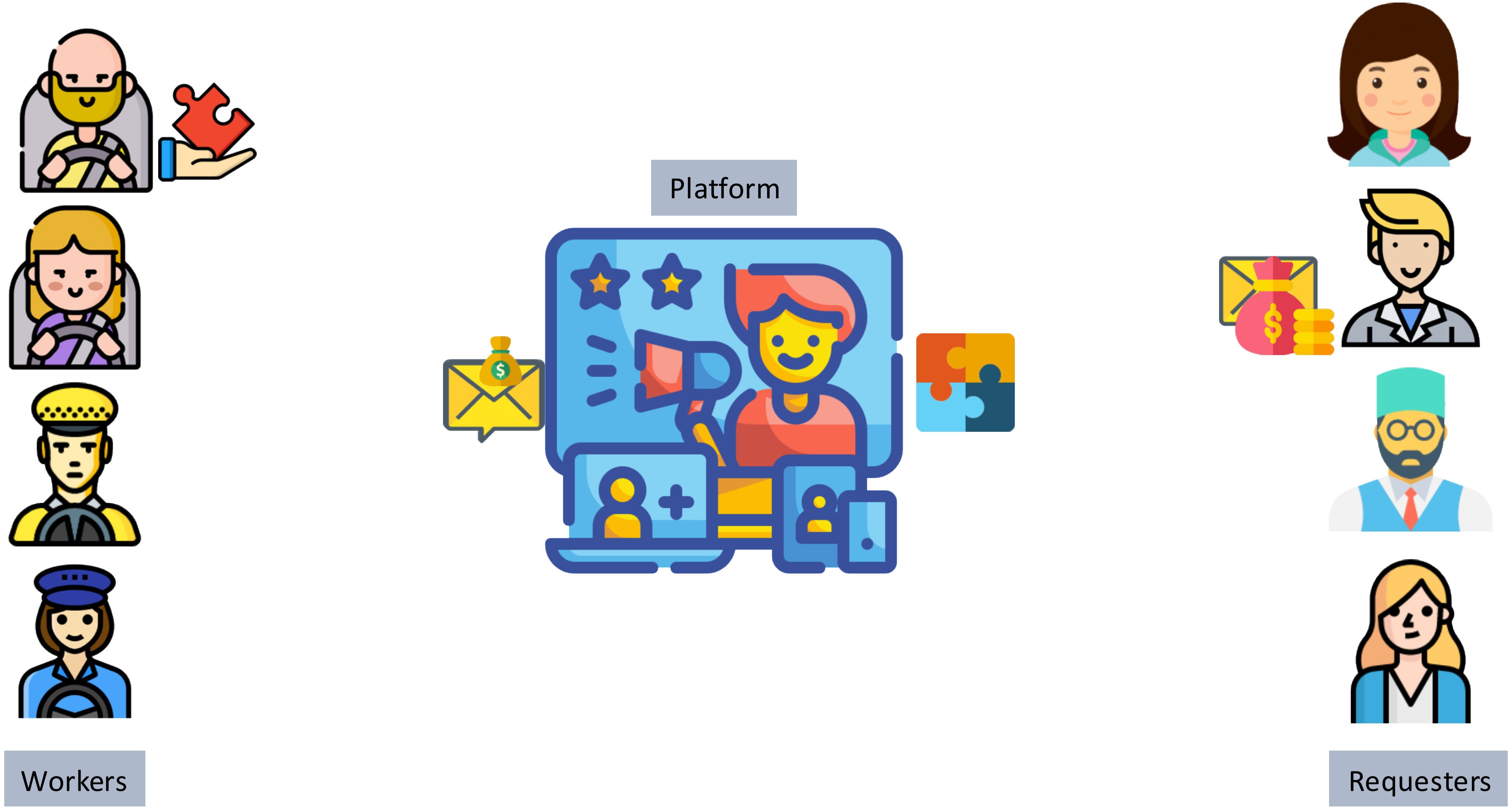
Privacy Meets Regulations: Shaping the Future of Work

Mohammad Javad Amiri[1], Tristan Allard[2], Boon Thau Loo[3],
Divyakant Agrawal[4], Amr El Abbadi[4]

[1]Stony Brook University, [2]University of Rennes, [3]University of Pennsylvania, [4]UC Santa Barbara

CIDR
Conference on Innovative Data Systems Research

Crowdworking Environments



Platform

Workers

Requesters

Guaranteeing the compliance of crowdworking platforms with regulations

> “Whereas universal and lasting peace can be established only if it is based upon social justice; . . . for example, by the regulation of the hours of work . . . ”

preamble of the constitution of the International Labor Organization
[Commission on International Labor Legislation, 1919]

Figure: Members of the Commission on International Labor Legislation to the Paris Peace Conference (1919).

FLSA: Total work hours of a worker per week may not exceed 40 hours

In California, Assembly Bill 5 (AB5) entitles workers to greater labor protections, such as minimum wage laws, sick leave, and unemployment and workers' compensation benefits.

CA Proposition 22 imposes its set of regulations, e.g., requires a worker to work at least 25 hours per week to qualify for healthcare subsidies.

Local regulations exist but not sufficient

- maximum driving time per day on Uber/Lyft

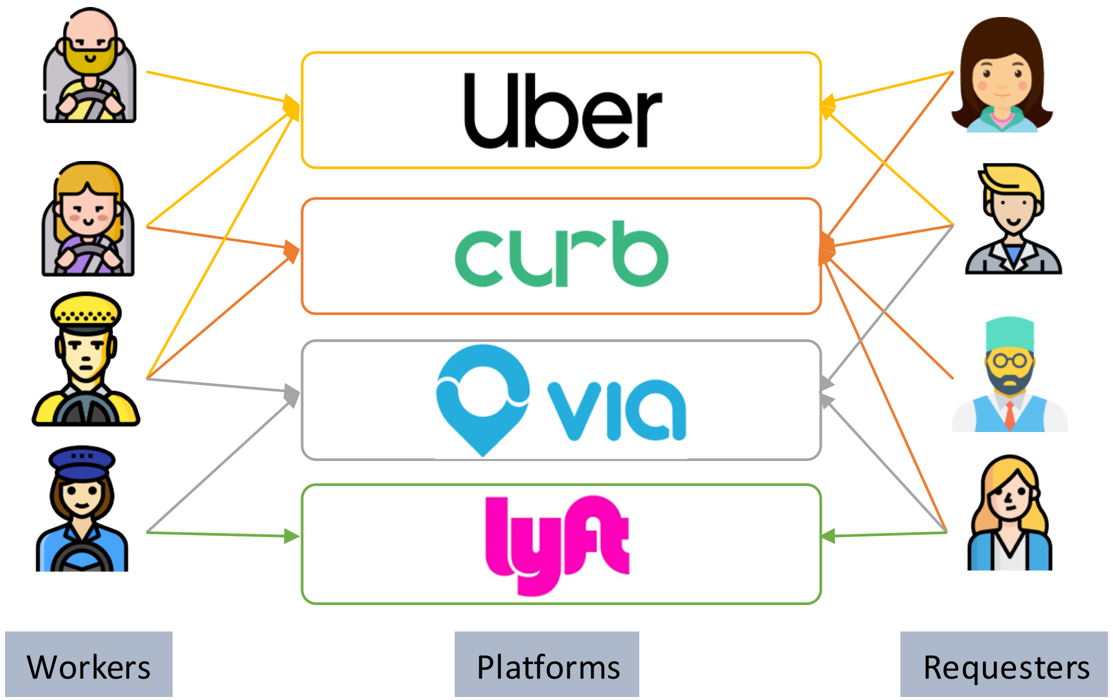
The Fair Labor Standards Act

was signed by President Franklin D. Roosevelt on June 25, 1938.

VOTE NO ON PROP. 22
PROP.22: SUPPORT GIG DRIVERS AND FOOD DELIVERY WORKERS
STOP EXPLOITATION OF GIG WORKERS Vote NO
ATU VOTE NO PROP 22
VOTE NO ON PROP 22: Bad for Workers, Bad for California

There is more than one platform …

- Workers often work on several platforms
  - Mircotasking: Amazon Mechanical Turk, Prolific, Clickworker, Toloka, Microworkers, Remotasks
- Requesters submit tasks on multiple platforms
- Complex tasks might require multiple contributions



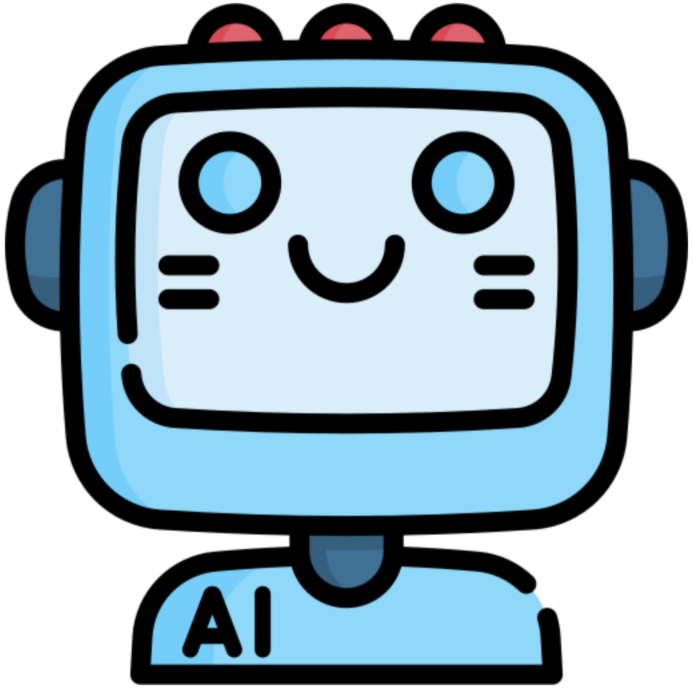
Privacy rights of participants

- No participant obtains or infers any information beyond what is strictly needed
  - A driver who works for both Uber and Lyft, does not want either of them know that she works for the other
- Reconcile transparency with privacy

Content of contributions

Use of LLMs enhances efficiency, accessibility and productivity. But

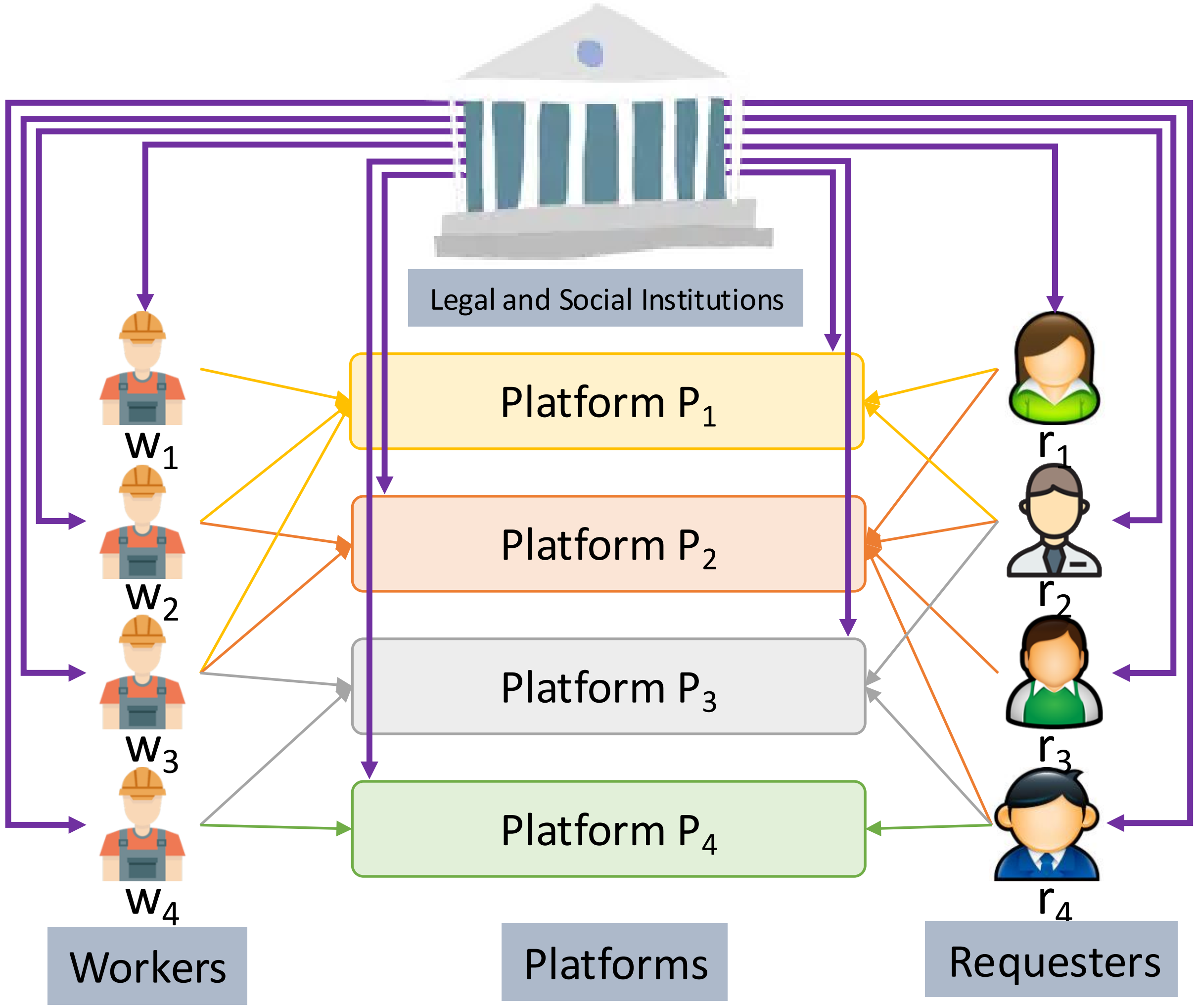
- Quality assurance
    - e.g., the quality and accuracy of generated content
- Ethical considerations
    - e.g., concerns over authorship, intellectual property rights, and the potential for generating misleading or harmful content
- Biases
    - e.g., resulting from inherit biases present in their training data
- Data privacy
    - e.g., violating user data confidentiality



Enforce global regulations on multi-platform crowdworking environments while preserving privacy

A vision for future regulation systems

- Four main design dimensions
  - D1: Type of supported regulations
    - e.g., aggregate or not
  - D2: Privacy guarantees
  - D3: Architecture of the system
    - e.g., fully decentralized
  - D4: Content validation mechanisms



Design space: D1. regulation type

- Express regulations as SQL constraints over a universal table U−TABLE
- Categorize them according to their SQL expression
- Characterized by:
    - Complexity: simple if no JOIN operation, complex otherwise
    - Aggregate (presence of Aggregate function(s), with GROUP BY and HAVING):
    - row-only, aggregate-only and mixed

- Enforceable: must always hold
    - e.g., maximum work hours
- Verifiable: must hold periodically
    - e.g., minimum work hours

WORKER	PLATFORM	REQUESTER	TIMECOST	CONTRIB
w1	p2	r1	3H	...
w1	p3	r2	2H	...
w2	p1	r1	6H	...

Design space: D2. privacy guarantees

- Threat model:
  - System dependent
    - e.g., honest-but-curious, covert, malicious

- Privacy model: pluggable disclosures (to be personalized):
  - Disclosures to the participants that are not involved in the the crowdworking process π and that have not received task t from requester r: $\delta^{\pi}_{\neg R \neg I}$
  - Disclosures to the platforms and workers that have received the task t from r but that are not involved in π: $\delta^{\pi}_{R \neg I}$
  - Disclosures to the participants that are directly involved in π (and have thus received task t): δ^{π}_{RI}

Design space: D3. architectural choices

- Main components:
  - Regulation management: models and enforces the regulations
  - Global state management: maintains the global state of the system

- Design and implementation choices:
  - Centralized
    - easier to rapid prototype
    - difficult to ensure fault-tolerance, privacy, and trustworthiness
  - Decentralized
    - more compatible with the multi-platform settings
    - resulting in more overhead and complex communication protocols among entities

A possible point in the design space

- Regulation supported:
  - `U-TABLE` focuses on the interactions and consists in
    - `WORKER`, `PLATFORM`, `REQUESTER`, `TIMECOST`
  - `Simple`, `mixed` with `SUM-aggregate` regulation, with lower-than (enforceable) or higher-than (verifiable) thresholds

A possible point in the design space

Privacy guarantees:

- Covert non-colluding adversaries
  - Aims at inferring anything that can be inferred from the execution sequence
  - Is able to deviate from the protocol if no other participant detects it
- Disclosures sets: (given crowdworking process π: (BEGIN, END, w, p, r, t))
  - $\delta^{\pi}_{\neg R \neg I}$ = (BEGIN, END, p)
  - $\delta^{\pi}_{R \neg I}$ = (BEGIN, END, p, r, t)
  - δ^{π}_{RI} = (BEGIN, END, w, p, r, t)

A possible point in the design space

Hybrid architecture:

- Centralized Registration Authority (RA)
    - Registers participants, models regulations, distributes crypto material
- Decentralized Multi-Platform Infrastructure
    - Maintains the global state within a replicated datastore
- Consensus protocols:
    - Local: across nodes of the same platform
    - Cross-platform: across platforms having received the same task
    - Global: across all platforms

A Simple token-based approach

- Implement enforceable and verifiable regulations by managing two budgets per participant
- Lightweight, single-use, and anonymous tokens

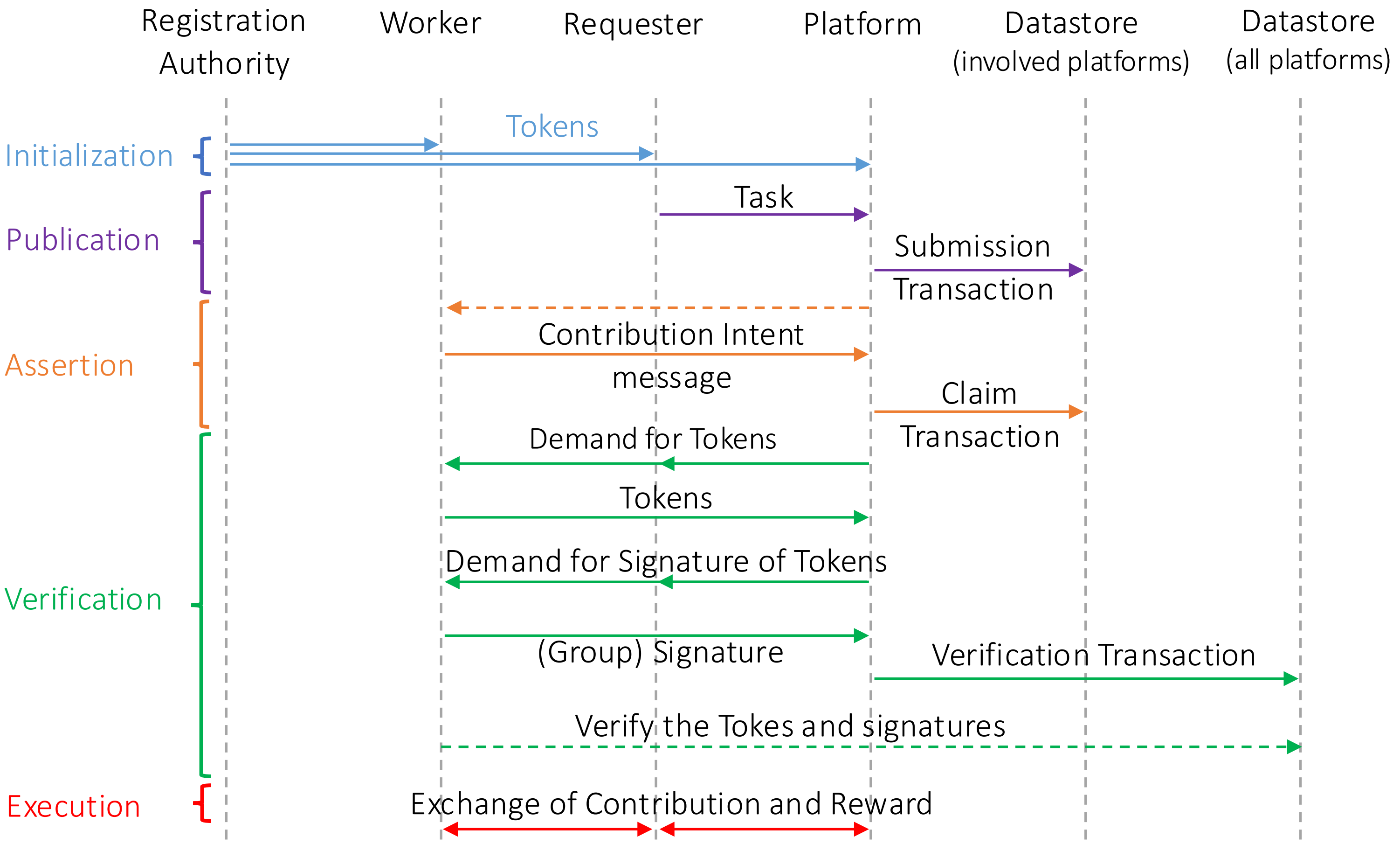
The registration authority refreshes participants tokens periodically

- GENERATE: initializing the budgets and refilling them
    - enforceable and verifiable tokens
- SPEND: spending portions of the budgets
- PROVE: providing proof for verifiable regulations to a third party
- CHECK: checking whether a given spending is allowed or not
- ALERT: reporting dubious spending

Execution sequence

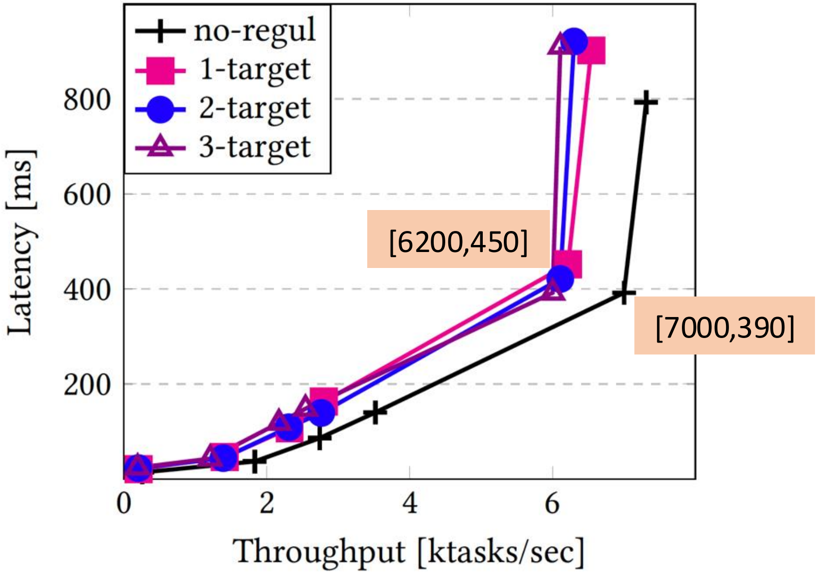
Tasks:
Internal
Cross-Platform

Transactions:
Submission
Claim
Verification



Different types of constraints

((w, *, *), θ)
((w, *, r), θ)
((w, p, r), θ)



[6200,450]

[7000,390]

Four platforms

Two constraints

10% cross-platform

Single contribution

Privacy-preserving mechanism: only 11% throughput and 15% latency overhead

The class of regulations does not significantly affect the performance

Conclusion

- An overall vision for future of work multi-platform regulation systems
- A simple token-based system to address the problem of enforcing global regulation over multi-crowdworking platforms
- How to (a) support other types of regulation, (b) provide different privacy guarantees or (c) eliminate the centralized RA?

Questions?

amiri@cs.stonybrook.edu