Optimally Auditing Adversarial Agents

Sanmay Das[1], Fang-Yi Yu[2], Yuang Zhang[2]
Virginia Tech, George Mason University

VIRGINIA TECH

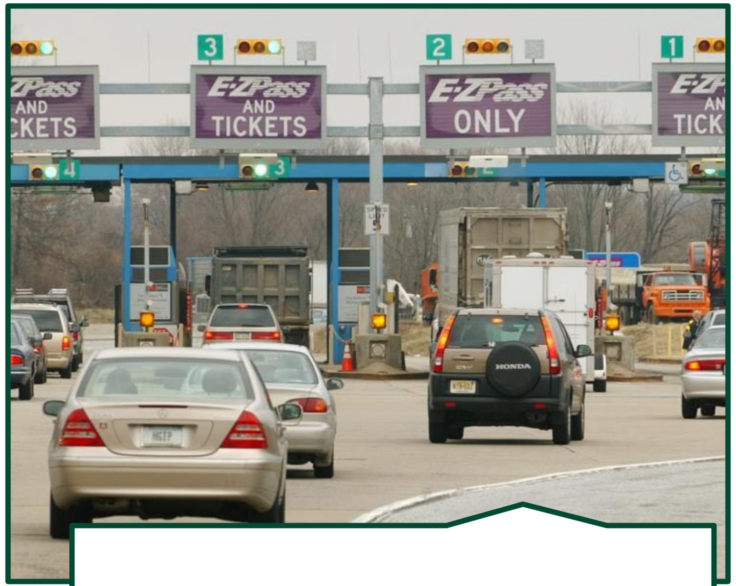
GEORGE MASON UNIVERSITY

Auditing in High-Stakes Domains

Society relies on self-reported data to allocate resources

Social Services & Government Benefits

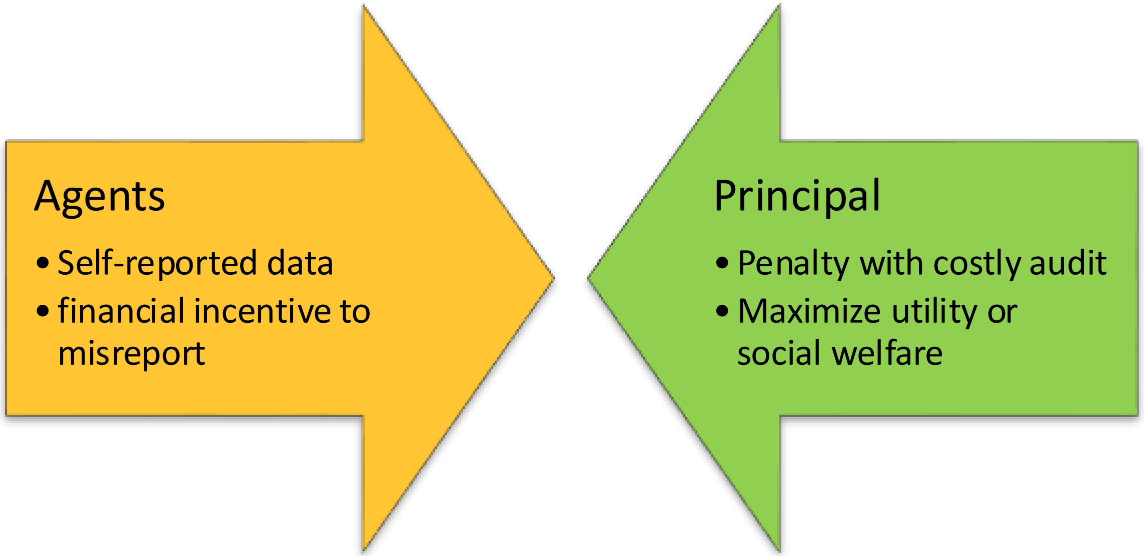
Tax Relief & Fraud



Toll Evasion

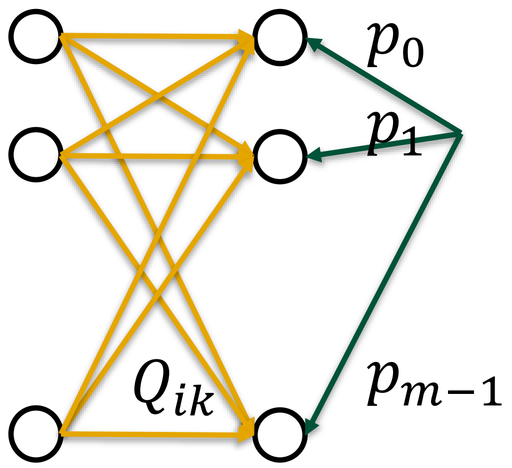
Research Problem

- The conflict
  - Agents: incentive to misreport (fraud)
  - Principal: verifying (auditing) is costly.
- Design an **audit strategy** against strategic coordination.



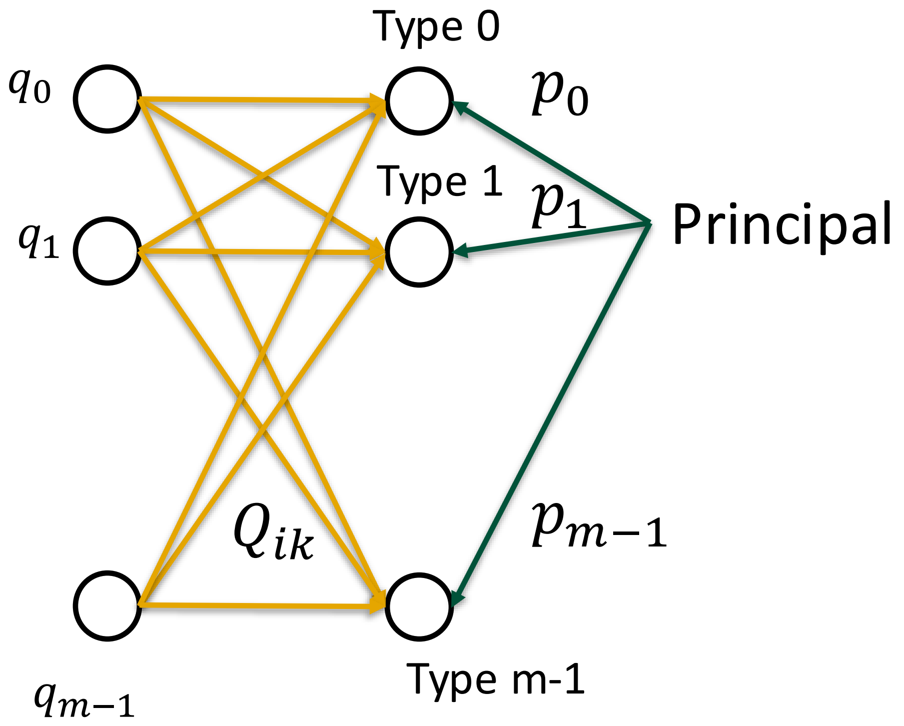
Model: Principal-Multi-agent Game

- Principal commits to an **audit vector** $p \in [0,1]^m$ on m types (e.g., level of income)
- Agents misreport their types under some equilibrium Q
- Payoff structure:
  - – Principal: agents' misreports(-), audit cost (-), penalty(+)
  - – Agents: misreport (+) and penalty(-)



Model: Principal-Multi-agent Game

- Principal commits to $p \in [0,1]^m$ with cost λ
- Each agent observes the private type $i \sim q$ and chooses $Q \in [0,1]^{m \times m}$
- Payoffs
  - Agent: $U_{ik} = pay(k) - p_k pen(i,k)$
  - Principal $V(p,Q) = \sum_{i,k} q_i Q_{ik}(val(i,k) - pay(k) + p_k(-\lambda + pen(i,k)))$
  - Affine penalty: $pen(i,k) = (pay(k) + b)1[i \neq k]$

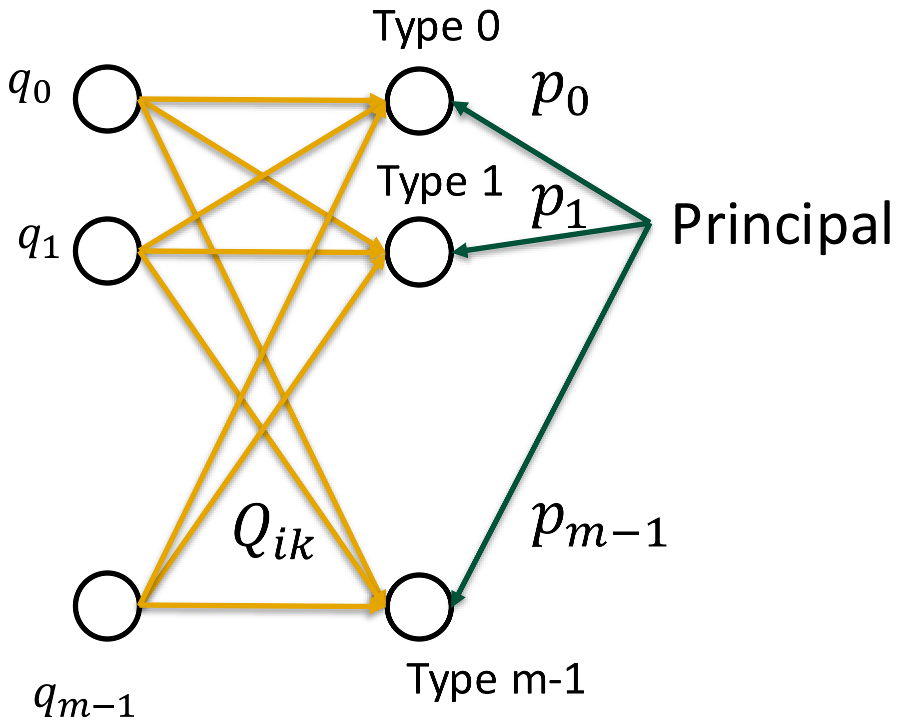


Model: Principal-Multi-agent Game

- Principal commits to p
- Agents choose Q
- Payoffs:
  - Principal: $V(p, Q)$
  - Agent: $U_{ik} = pay(k) - p_k pen(i,k)$
- **Bayes-Nash Equilibrium**:

$$U_{ik} \geq U_{il},$$

for all k, l with $Q_{ik} > 0$.



Goal and Challenge

- Goal: find the optimal audit vector to maximize the principal's utility when agents play the worst equilibrium

$$\max_{p} \min_{Q \in Eqi(p)} V(p, Q)$$

  - Multiple possible equilibria
  - Large non-convex variable spaces: $p \in [0,1]^m$ and $Q \in [0,1]^{m \times m}$

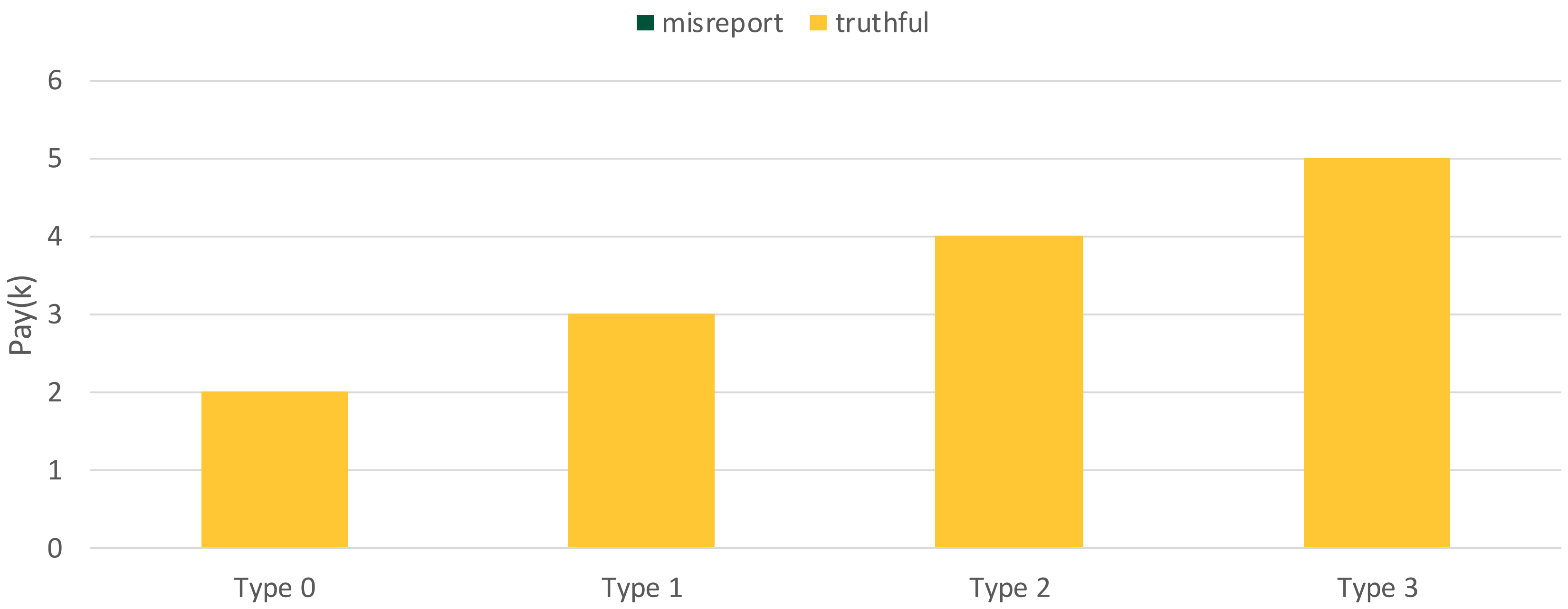
Optimizing the Principal's Utility

Theorem 1 (Utility-optimal)**.** *For any small enough* $\epsilon > 0$, $(n, m, \boldsymbol{q}, \mathrm{val}, \mathrm{pay}, \mathrm{pen})$ *and* λ*, Algorithm 1 computes a* $2n\epsilon$*-optimal audit vector for Eq. (7) in* $O(m^4)$ *time.*

Moreover, the time complexity can be improved to $O(m^2)$*.*

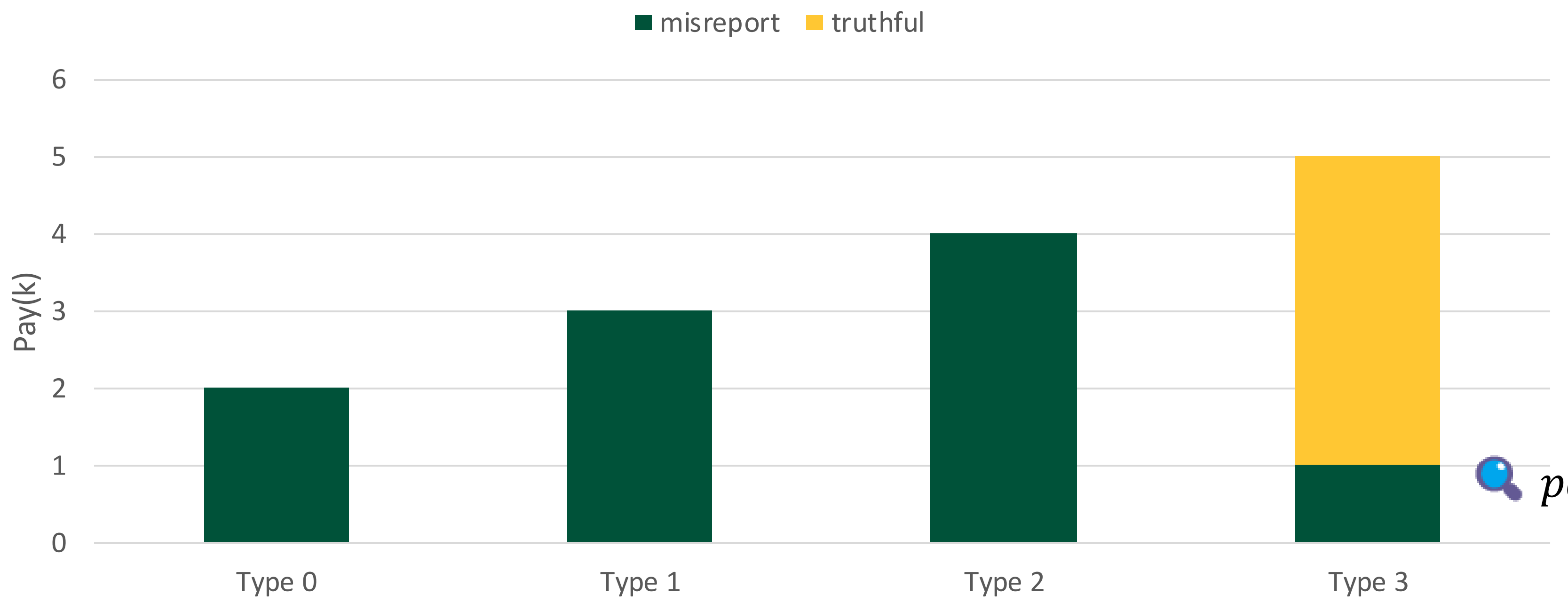
Agents' Best Response

- Idea: agents' best response is a threshold strategy



Agents' Best Response

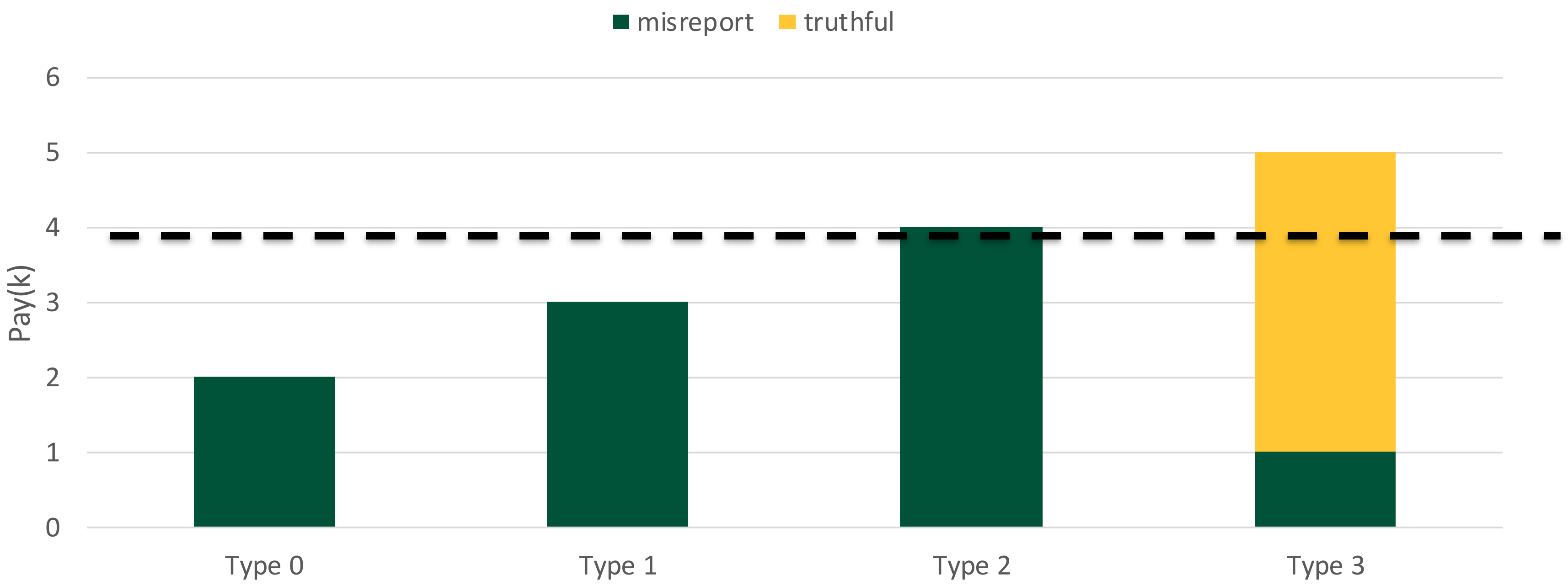
- Audit the highest type (type 3)



$pay(k) - p_k pen(k)$

Agents' Best Response

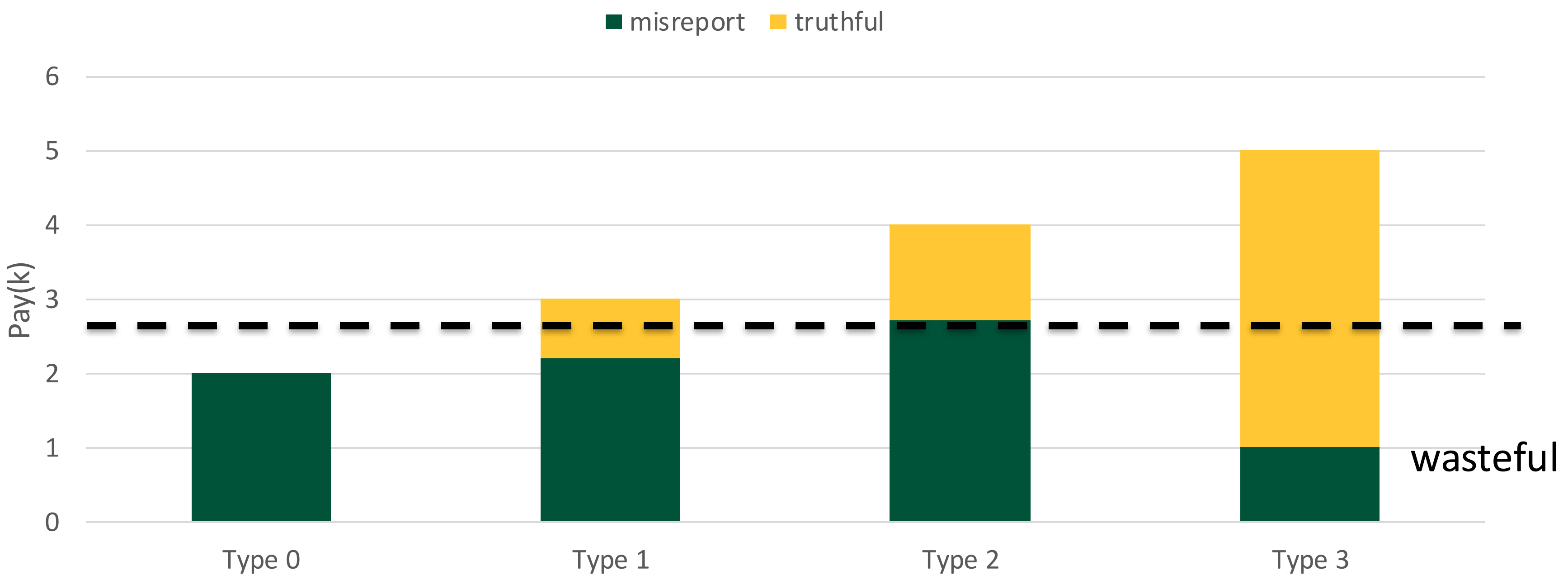
- Audit the highest type (type 3)
  - Everyone misreports to type 2



Equilibrium Analysis

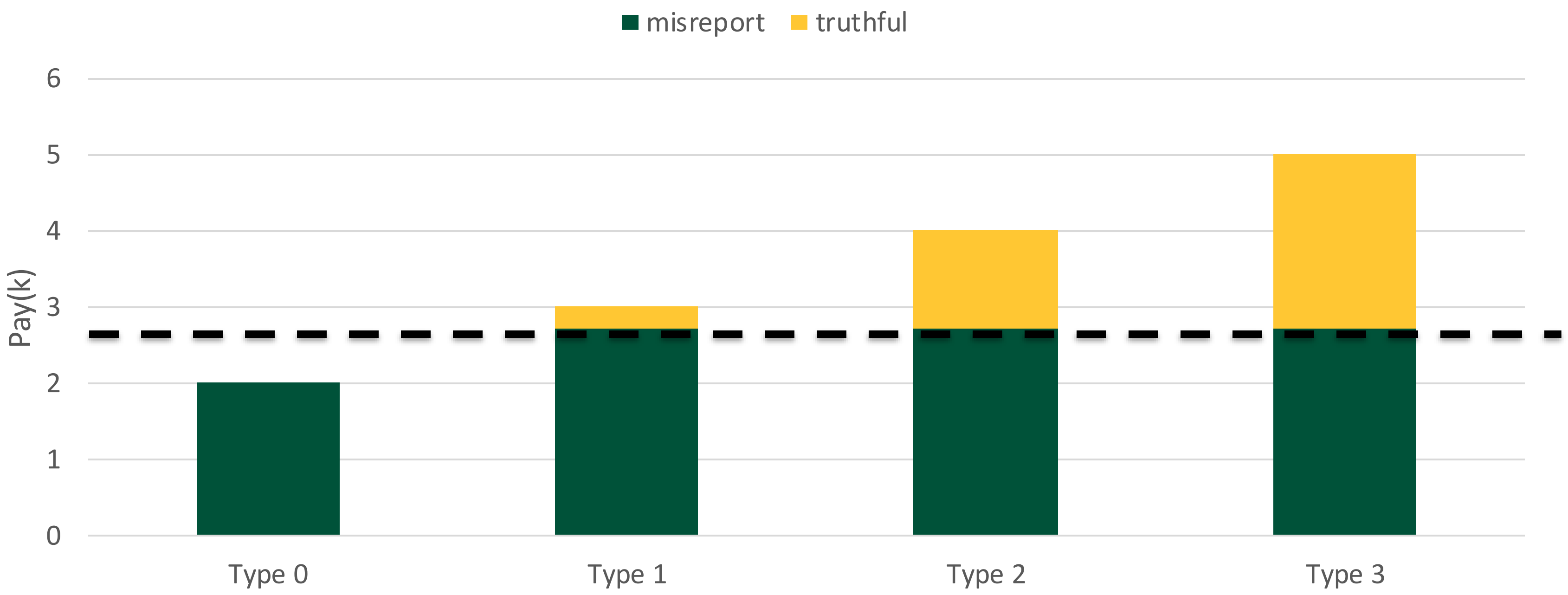
- Given any audit vector, misreport to k^* with the largest misreport payment

$$\widehat{U}_k = pay(k) - p_k pen(k)$$



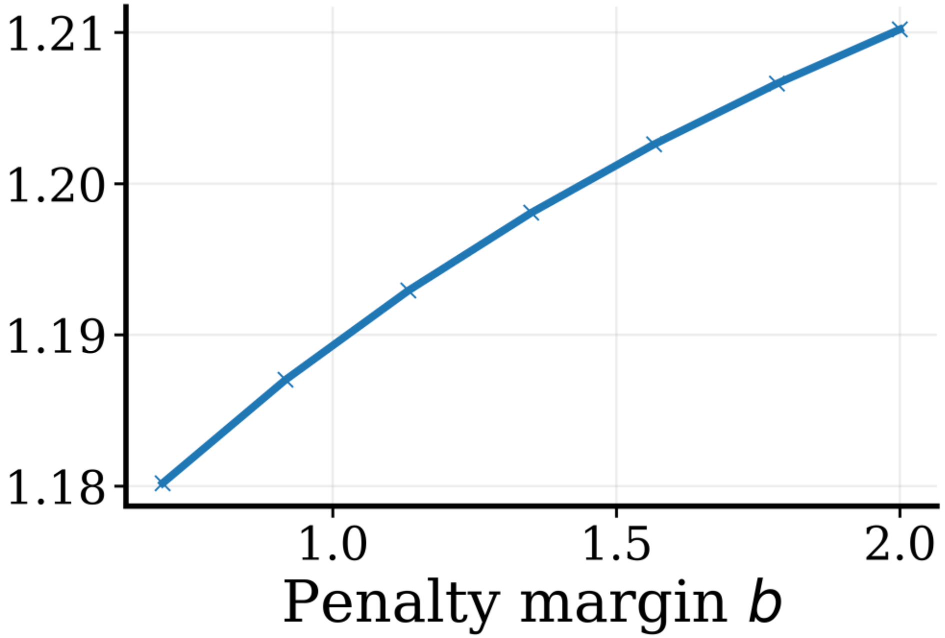
Equalized/Critical Audit Vector

- Choose the audit vector that equalizes $\widehat{U}_k$
  - No wasteful audits
  - Reduce the variable space

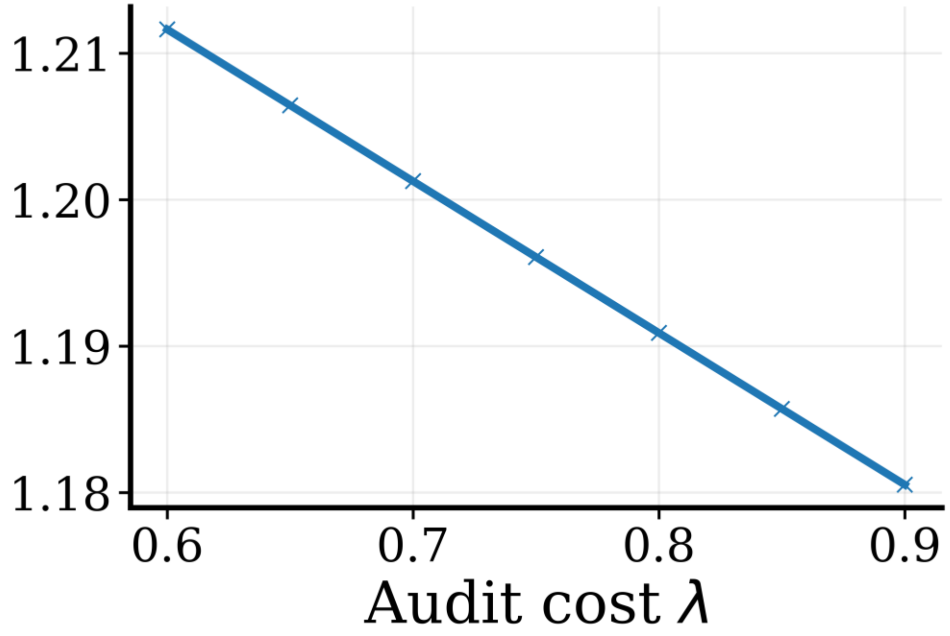


Monotone Impact of Audit Cost and Penalty

- Increasing penalties multiplies audit power



- Decreasing audit cost improves viability



Extensions

Unknown prior q

- no-regret algorithms (EXP3) on critical audit vectors

Adaptive audit strategy

- principal chooses a function $\pi: \Delta_m \to [0,1]^m$ outputting audit vectors
- Adaptive = non-adaptive if $pen(i,k) = (pay(k) + b)1[i \neq k]$

Conclusion

- Summary
  - Modeling auditing as a pessimistic Stackelberg game
  - Optimal approximation algorithm for utility and welfare
  - Monotone impact of audit cost and penalty
  - Variants: unknown parameter and adaptive strategy
- Future work
  - Generalize to finite agents, noisy or partial verification, and richer penalty structures.
  - Design problem of payment and penalty function.