

I8739A: Foundations of Security and Privacy

## Regret Minimizing Audits

Anupam Datta  
Carnegie Mellon University

# Example from HIPAA Privacy Rule

A covered entity may disclose an individual's protected health information (phi) to law-enforcement officials for the purpose of identifying an individual if the individual made a statement admitting participating in a violent crime that the covered entity believes may have caused serious physical harm to the victim

## ▶ Concepts in privacy policies

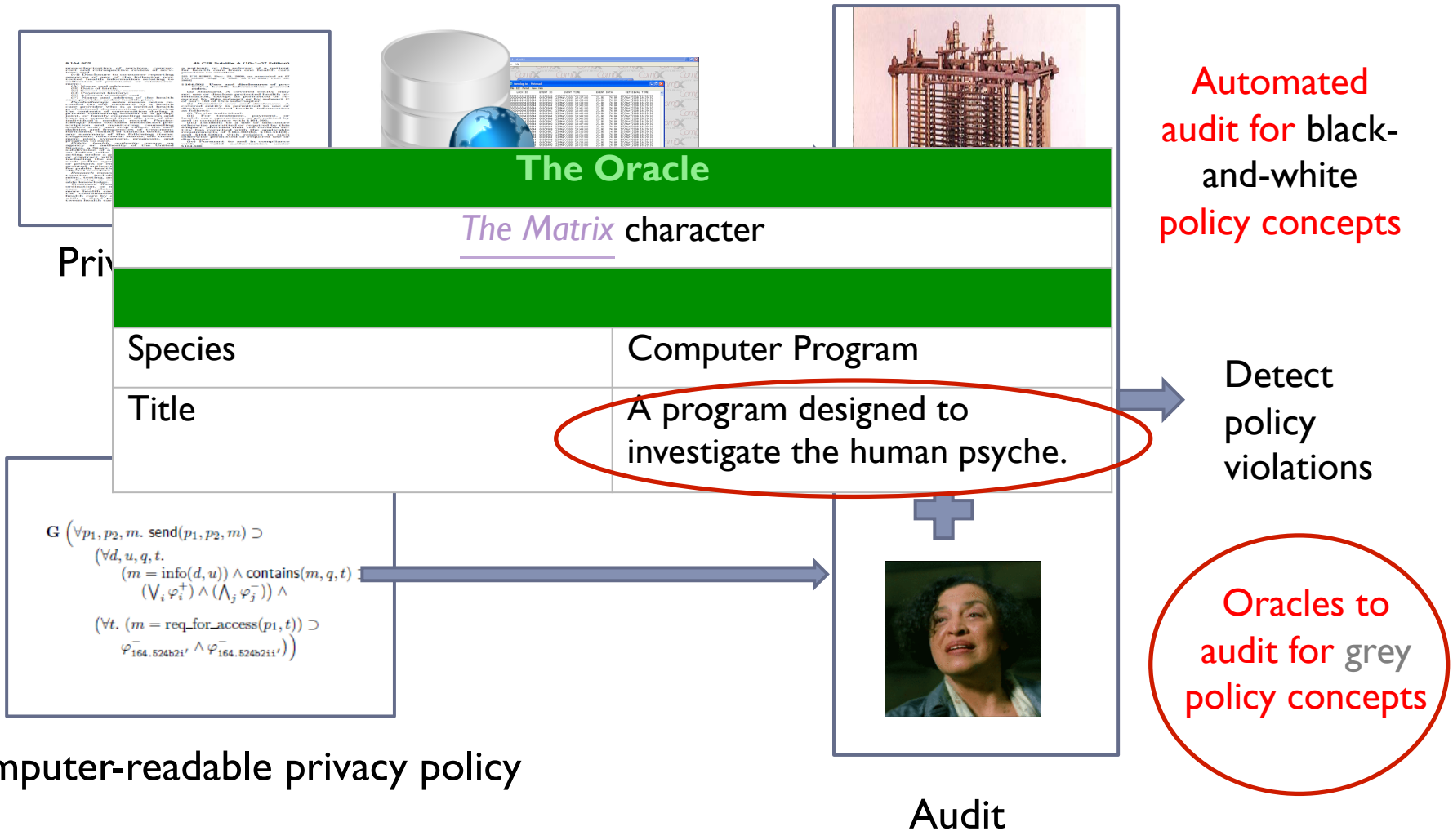
- ▶ **Actions:** send(p1, p2, m)
- ▶ **Roles:** inrole(p2, law-enforcement)
- ▶ **Data attributes:** attr\_in(prescription, phi)
- ▶ **Temporal constraints:** in-the-past(state(q, m))
  
- ▶ **Purposes:** purp\_in(u, id-criminal))
- ▶ **Beliefs:** believes-crime-caused-serious-harm(p, q, m)

Black-and-white concepts

Grey concepts



# Audit Mechanisms for Privacy Protection



## Next Two Lectures

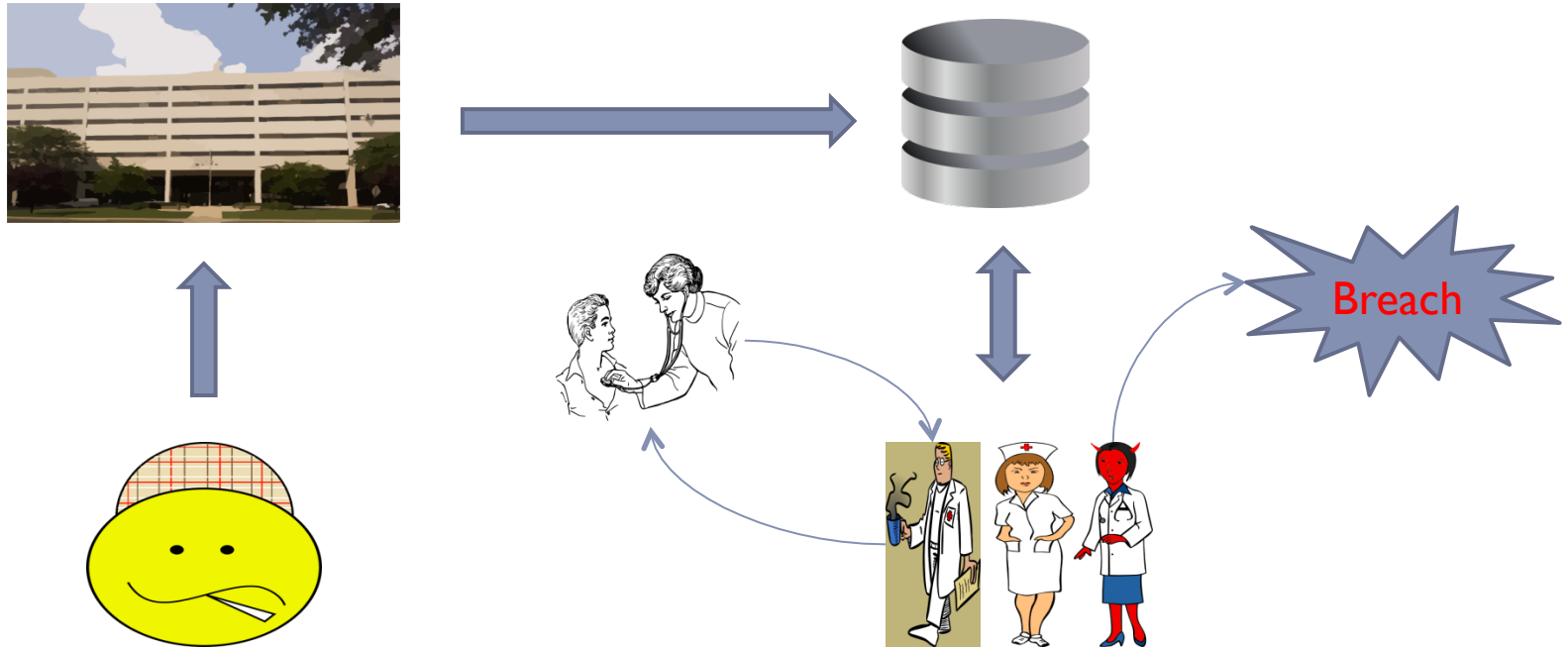
---

- ▶ Learning algorithm (oracle) to audit “grey” concepts
- ▶ Formalizing and enforcing *purpose* restrictions
  - ▶ Implementing oracle for purpose restrictions



# Motivation

---



- ▶ Goal: treatment
- ▶ Rigid access control hinders treatment
- ▶ Permissive access control  $\Rightarrow$  human in the loop  $\Rightarrow$  privacy violations

# Enforcement Using Audits

---

- ▶ **Permissive access control**
  - ▶ If in doubt allow access
- ▶ **Log the accesses**
- ▶ **Review the accesses later and find violations**
  
- ▶ **Adhoc approaches in practice**
  - ▶ FairWarning audit tool implements simple heuristics, e.g., flag all celebrity access

# Humans are the Weakest Link

---



- ▶ Human actors trusted to act responsibly (e.g., use personal health information for treatment)
- ▶ Goal: Auditing human actors to ensure that they act responsibly

# Desiderata

---

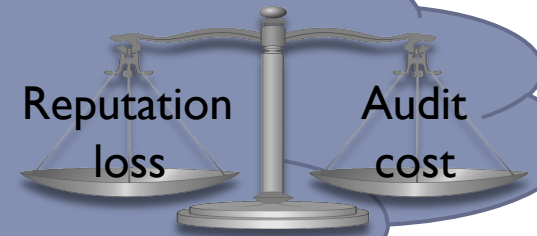
- ▶ **Principled study of the audit process**
  - ▶ A model (including humans) for audit process
  - ▶ Properties of the audit mechanism
  - ▶ Audit mechanism which provably satisfies the property

# Audit Algorithm by Example



Auditor

Auditing budget: \$3000/ cycle  
 Cost for one inspection: \$100  
 Only 30 inspections per cycle  
**Employee incentives unknown**



Access divided into 2 types

Reputation Loss from 1 violation (internal, external)

100 accesses

30 accesses



Sandra Bullock

**\$500, \$1000**

70 accesses



**\$250, \$500**



# Audit Algorithm Choices



Only 30 inspections

Consider 4 possible allocations  
of the available 30 inspections



Sandra Bullock



Weights

|     |     |     |     |
|-----|-----|-----|-----|
| 0   | 10  | 20  | 30  |
| 30  | 20  | 10  | 0   |
| 1.0 | 1.0 | 1.0 | 1.0 |

Choose allocation probabilistically based on weights

# Audit Algorithm Run



| No. of Access | Actual Violation |
|---------------|------------------|
| 30            | 2                |
| 70            | 4                |



|    |    |    |    |
|----|----|----|----|
| 0  | 10 | 20 | 30 |
| 30 | 20 | 10 | 0  |

Observed

Estimated

Loss

Loss

| Int. Caught | Ext. Caught |
|-------------|-------------|
| 1           | 1           |
| 2           | 1           |



|        |        |        |        |
|--------|--------|--------|--------|
| \$2000 | \$1500 | \$1000 | \$1000 |
| \$750  | \$1250 | \$1250 | \$1500 |

Updated weights

|     |     |     |     |
|-----|-----|-----|-----|
| 0.5 | 0.5 | 2.0 | 1.5 |
|-----|-----|-----|-----|

Learn from experience: weights updated using observed and estimated loss

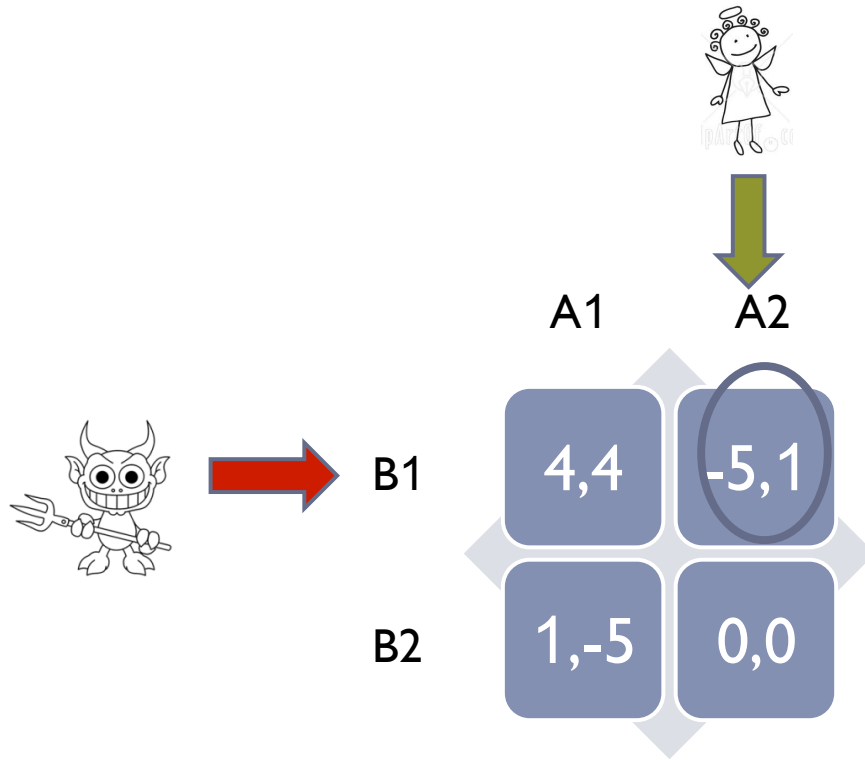
## Main Contributions

---

- ▶ A game model for the audit process
- ▶ Defining a desirable property of audit mechanisms, namely *low regret*
- ▶ An efficient audit mechanism RMA that provably achieves *low regret*

# A One-Shot Game Model

---



Strategic interaction:

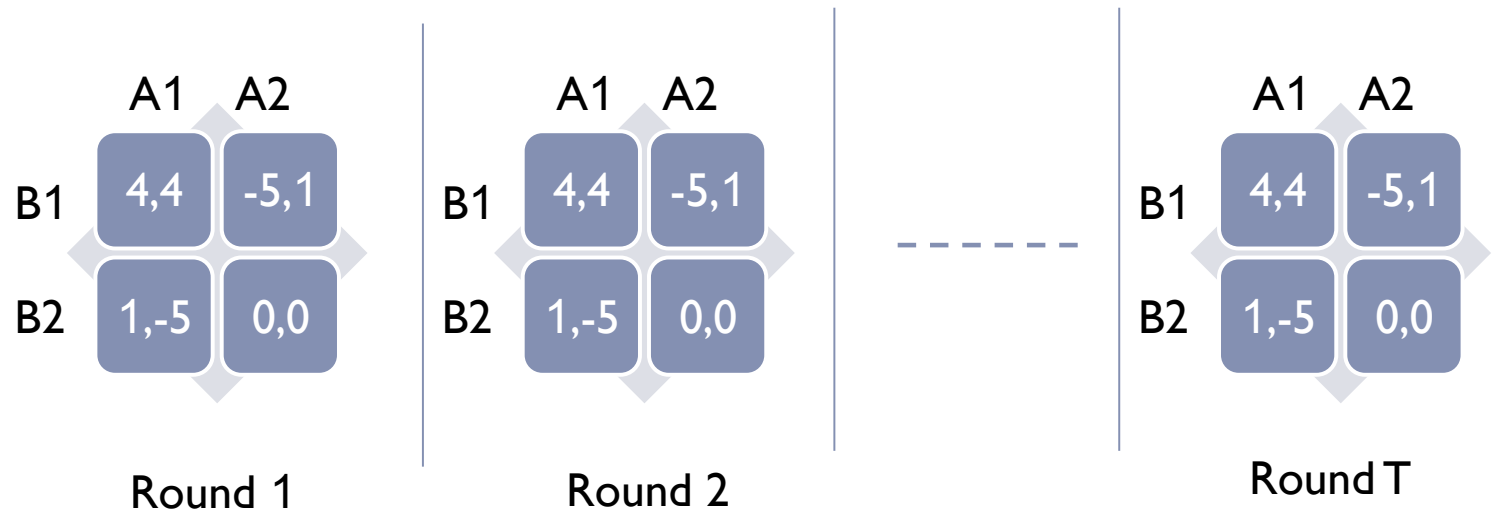
A player's reward depends on actions of other players

---



# A Repeated Game Model

---

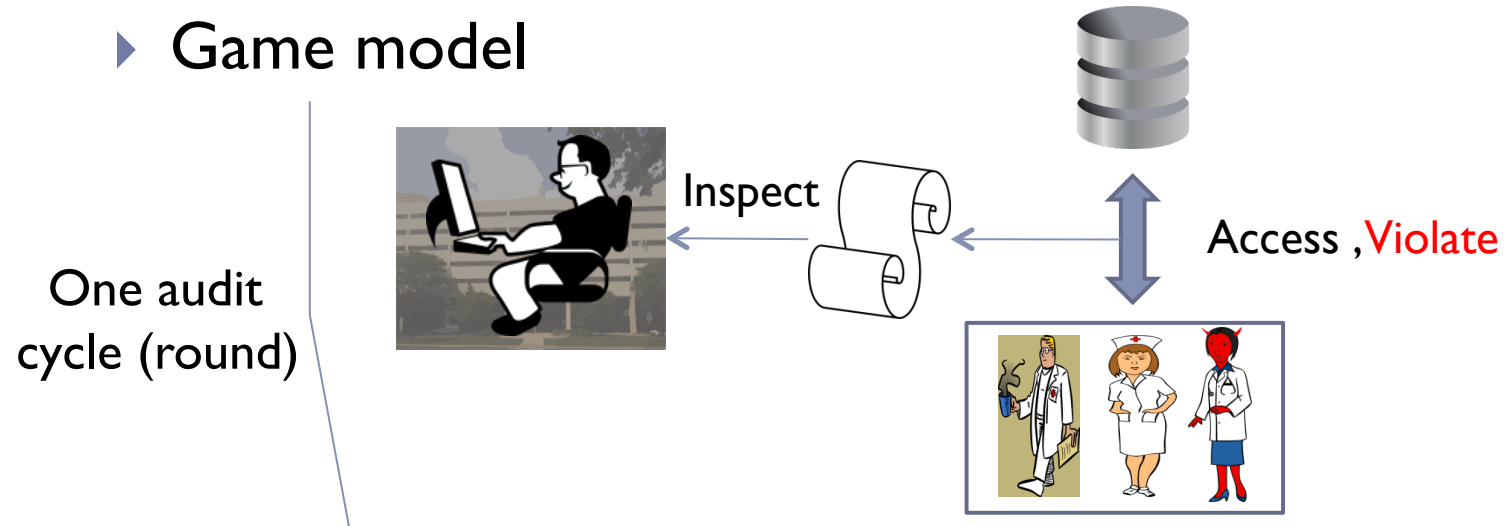


- ▶ Players interact repeatedly



# Repeated Game Model for Audit

## ▶ Game model



## ▶ Typical actions in one round

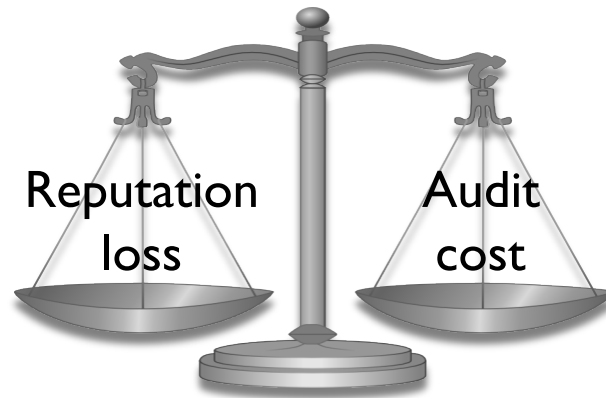
- ▶ Emp action: (access, **violate**) =  $([30,70], [2,4])$
- ▶ Org action: inspection =  $([10,20])$

Imperfection

# Game Payoffs

---

- ▶ Organization's payoff



- ▶ Audit cost depends on the number of inspections
- ▶ Reputation loss depends on the number of violations caught

- ▶ Employee's payoff unknown

# Auditing Model: Payoff

- ▶ Fixed cost for inspection of each type of task

- ▶ Example:

- cost per inspection:  $\langle 4, 5 \rangle$  ,
    - budget = 1500,  $s = \langle x, y \rangle$

- $$\text{cost}(s) = 4x + 5y \leq 1500$$

- ▶ Payoff of organization

- ▶ Cost of audit

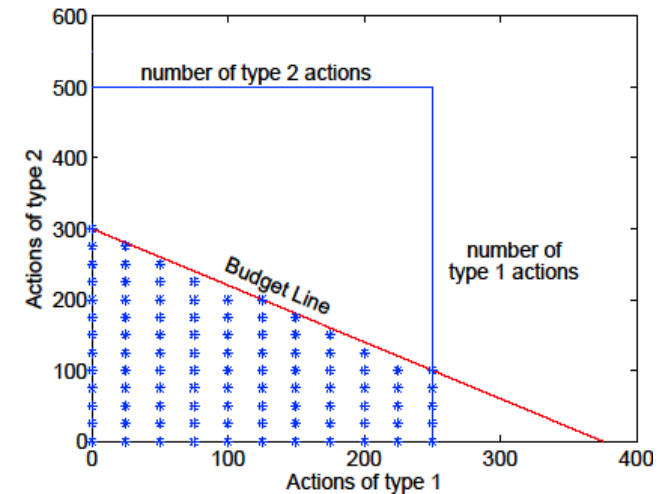
- ▶ Effect of detected violation on reputation

- ▶ Reputation  $R$  goes down with detected violation

- ▶ Payoff =  $-\text{cost}(s) + R$

- ▶ Payoff of employees

- ▶ We do not assume anything about the employee's incentives (worst-case adversary model, not rational)



# Auditing model: Reputation

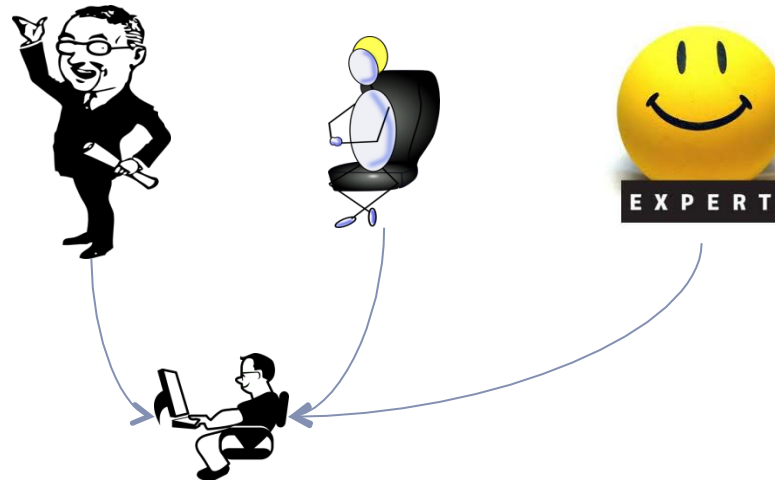
---

- ▶ Only detected violations in last  $m$  rounds affect reputation
- ▶ Recent violations cause more loss in reputation
- ▶ Example
  - $m = 2$ , one type of task, max. reputation  $Rep$  and detected violations in last two rounds are  $\langle 7 \rangle$  and  $\langle 4 \rangle$  respectively
  - ▶  $R = Rep - 7 - 4\delta$
- ▶ Payoff depends also on actions of previous rounds (not a repeated game)
- ▶ Accounting for future losses in the present
  - ▶ Suppose next violation is  $\langle 5 \rangle$ , then  $R = Rep - 5 - 7\delta$ .
  - ▶ The violation  $\langle 7 \rangle$  causes a total loss of  $7 + 7\delta$
  - ▶ Account for the effect of  $\langle 7 \rangle$  in the round when  $\langle 7 \rangle$  happened. Use  $R = Rep - 7 - 7\delta$
  - ▶ Now we have a repeated game model



# Property of Effective Audit Mechanism

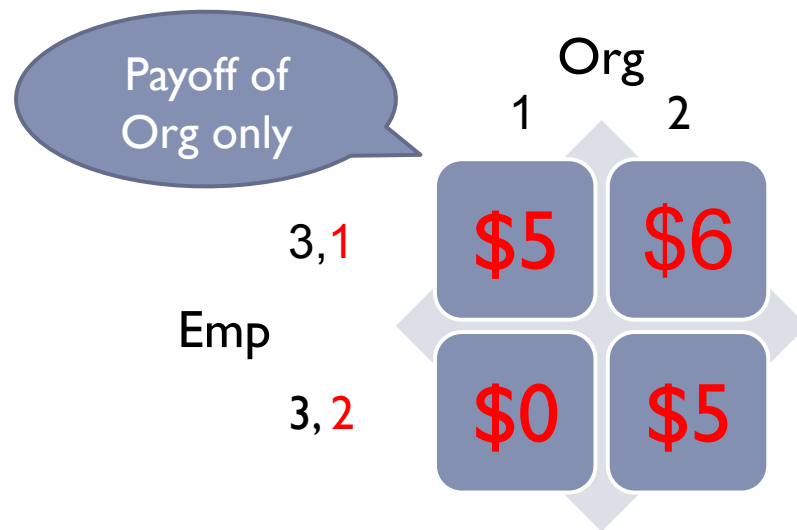
---



- ▶ Audit mechanism should be comparable to best expert in hindsight (concept of *regret* in online learning theory)
- 



# Regret by Example



Strategy: outputs an action for every round

$$\text{Total regret}(s, s1) = (-5) - (-6) = 1$$

$$\text{regret}(s, s1) = 1/2$$

| Players   | Round 1  | Round 2  | Total Payoff   |
|---|--|--|--|
| <ul style="list-style-type: none"> <li>Emp</li> <li>Org: s</li> </ul> | <ul style="list-style-type: none"> <li>3,1</li> <li>2 (\$6)</li> </ul> | <ul style="list-style-type: none"> <li>3,2</li> <li>1 (\$0)</li> </ul> | <ul style="list-style-type: none"> <li>Unknown</li> <li>\$6</li> </ul> |
| Org : s1  | 1 (\$5)  | 1 (\$0)  | \$5  |

## Meaning of Regret

---

- ▶ Low regret of  $s$  w.r.t.  $s^1$  means  $s$  performs as well as  $s^1$
- ▶ Desirable property of an audit mechanism
  - ▶ Low regret w.r.t all strategies in a given set of strategies
  - ▶  $regret \rightarrow 0$  as  $T \rightarrow \infty$

# Regret Minimization Algorithm

---

## Multiplicative weight update (MWU)

- ▶ Initially all strategies equally weighted:  
 $w_s = 1$  for all  $s$  in set of  $N$  strategies
- ▶ Probability distribution over strategies  $D_t(s) \propto w_s$
- ▶ Obtain the payoff vector  $Pay$ 
  - ▶ Payoff for all strategies known (perfect information)
- ▶ Update weights

$$w_s = w_s \cdot \gamma^{-Pay(s) + \gamma \cdot \sum_{s'} D_t(s') Pay(s')}$$

# Regret Bound for MWU Algorithm

---

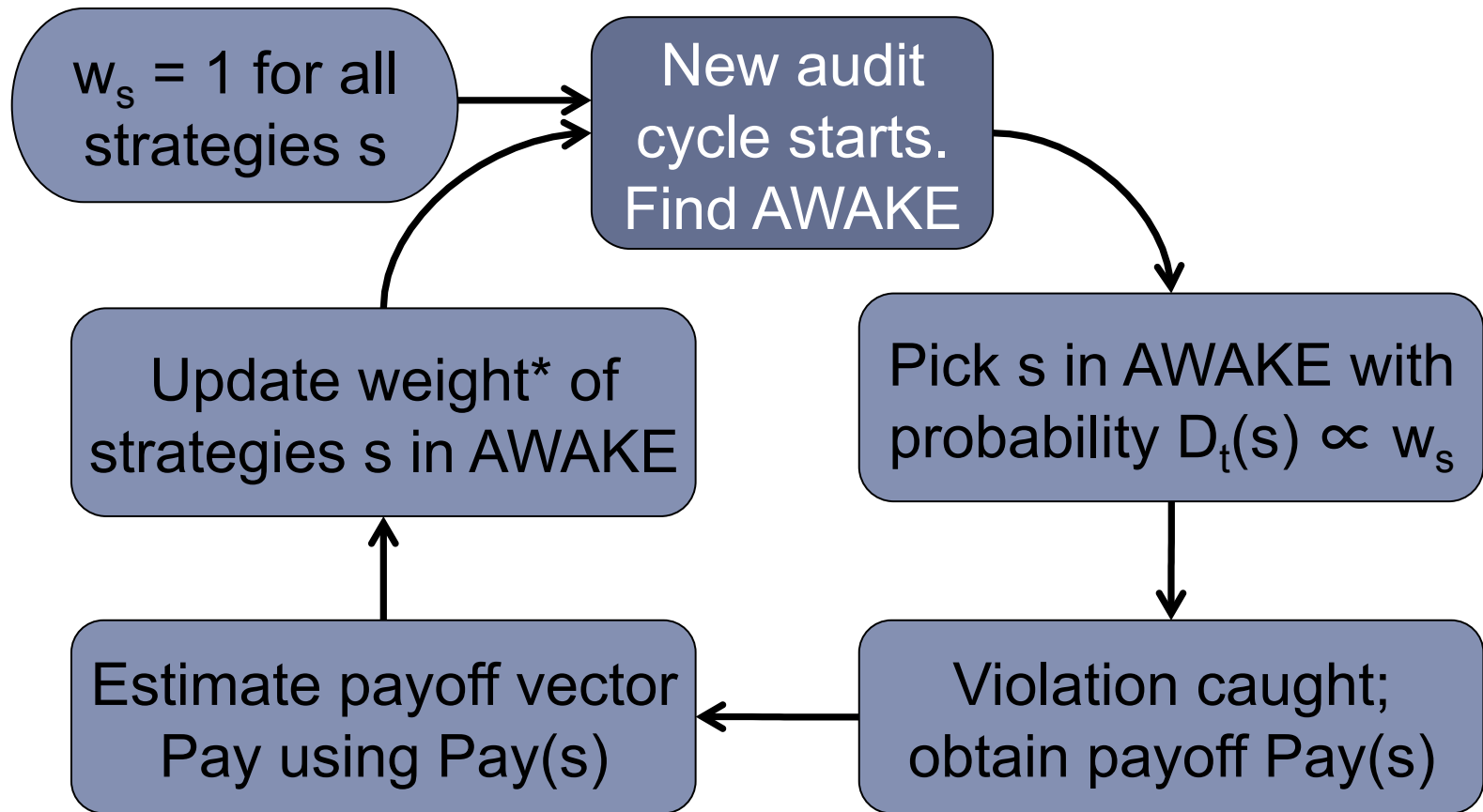
- ▶ The regret bound of MWU is

$$2\sqrt{\frac{\ln N}{T}}$$

- ▶  $N$ : number of strategies in the given set
  - ▶  $T$ : number of rounds of the game
  - ▶ All payoffs scaled to lie in  $[0,1]$
- 
- ▶ Why not MWU?
    - ▶ Imperfect information, unavailable strategies



# Regret Minimizing Audits (RMA)



$$*w_s \leftarrow w_s \cdot \gamma^{-\text{Pay}(s) + \gamma \cdot \sum_{s'} D_t(s') \text{Pay}(s')}$$

# Audit Algorithm Run



| No. of Access | Actual Violation |
|---------------|------------------|
| 30            | 2                |
| 70            | 4                |



|    |    |    |    |
|----|----|----|----|
| 0  | 10 | 20 | 30 |
| 30 | 20 | 10 | 0  |

Observed Loss      Estimated Loss

| Int. Caught | Ext. Caught |
|-------------|-------------|
| 1           | 1           |
| 2           | 1           |



|        |        |        |        |
|--------|--------|--------|--------|
| \$2000 | \$1500 | \$1000 | \$1000 |
| \$750  | \$1250 | \$1250 | \$1500 |

Updated weights

|     |     |     |     |
|-----|-----|-----|-----|
| 0.5 | 0.5 | 2.0 | 1.5 |
|-----|-----|-----|-----|

Learn from experience: weights updated using observed and estimated loss

## Guarantees of RMA

---

- ▶ With probability  $1 - \epsilon$  RMA achieves the regret bound

$$2\sqrt{\frac{2 \ln N}{T}} + \frac{2 \ln N}{T} + 2\sqrt{\frac{2 \ln \left(\frac{4N}{\epsilon}\right)}{T}}$$

- ▶  $N$  is the set of strategies
  - ▶  $T$  is the number of rounds
  - ▶ All payoffs scaled to lie in  $[0,1]$
- 
- ▶ Better bound than any existing algorithm

## Take Away Message

---

Optimize costs, given *imperfect* nature of periodic audits in a setting with *adaptive adversaries* whose *incentives* are not known

---

---



Receive important updates and special reports

▶ Subscribe to the Ponemon News Feed

Providing research and strategic consulting to the:

- > CIO and CISO
- > Privacy Leader
- > Marketer and Brand Manager

Home » Blog » Dr. Ponemon's blog » Cost of a data breach climbs higher »

## Cost of a data breach climbs higher



March 8, 2011

Most privacy advocates and people in the data protection community believe that data breach costs will start coming down eventually because consumers will become somewhat immune to data breach news. The idea is that data breach notifications will become so commonplace that customers just won't care anymore.

But, that hasn't happened yet. The latest [U.S. Cost of a Data Breach](#) report, which was just released today, shows that costs continue to rise. This year, they reached \$214 per compromised record and averaged \$7.2 million per data breach event. The fact is that individuals still care deeply about their personal information and they lose trust in companies that fail to protect it.

It's not only direct costs of a data breach, such as notification and legal defense costs that impact the bottom line for companies, but also indirect costs like lost customer business due to abnormal churn. This year's study showed some very interesting results. In my view, there are a few standout trends.

# Auditing Model: Detection of Violations

---

- ▶ **Internal detection: hyper-geometric distribution**
  - ▶ Probability of picking  $k$  red balls (detecting  $k$  violations) when  $s$  balls are picked ( $s$  actions are inspected) from a bag of  $a$  balls (total # of tasks) of which  $v$  are red (# violations) and the rest are blue
- ▶ **External detection: fixed probability  $p$** 
  - ▶ Observe: No incentive to audit if violations not detected externally
  - ▶ In practice: regulations, market forces
- ▶ **Probabilistic outcome in each round**
  - ▶ Not all violations are detected since not all tasks are inspected



# Challenge: Imperfect Information

---

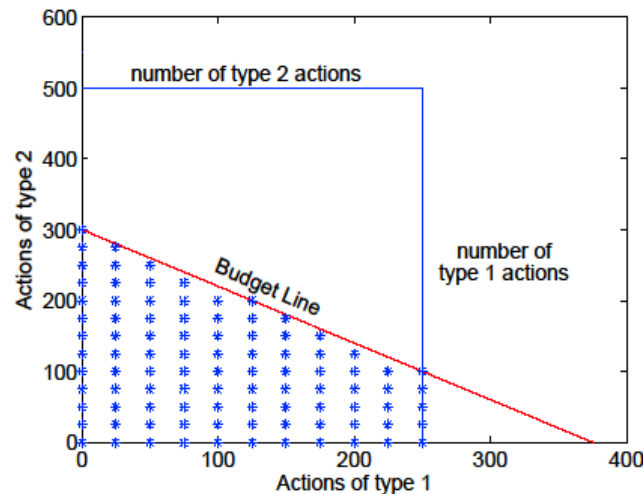
- ▶ Organization receives payoff corresponding to the action it played
- ▶ How to find the payoff for some other action?
- ▶ Our approach:
  - An estimator  $est(Pay(s), s')$ 
    - ▶ Error  $X(s', t) = est(Pay^t(s), s') - Pay^t(s')$
    - ▶ Use estimated payoffs
    - ▶ Bound difference between estimated and real payoff assuming certain properties of  $est$



# Challenge: Sleeping Experts

---

- ▶ Our set of experts includes all fixed inspections
- ▶ Sleeping experts [BM05]
  - ▶ Not all experts available in each round
- ▶ The organization's action space varies from round to round depending on which experts are sleeping



## Related Work

---

- ▶ **Weighted Majority Algorithm [LW89]:**
  - ▶ Average Regret:  $O((\log N)/T)^{1/2}$
  - ▶ Defender cannot run this algorithm unless he observes the adversaries moves (perfect information setting)
  
- ▶ **Imperfect Information Setting [ACFS02]:**
  - ▶ Average Regret:  $O(((N \log N)/T)^{1/2})$
  - ▶ Regret bound converges to 0 much slower
  
- ▶ Our regret bounds are of the same order as the perfect information setting



---

Thanks!  
Questions?