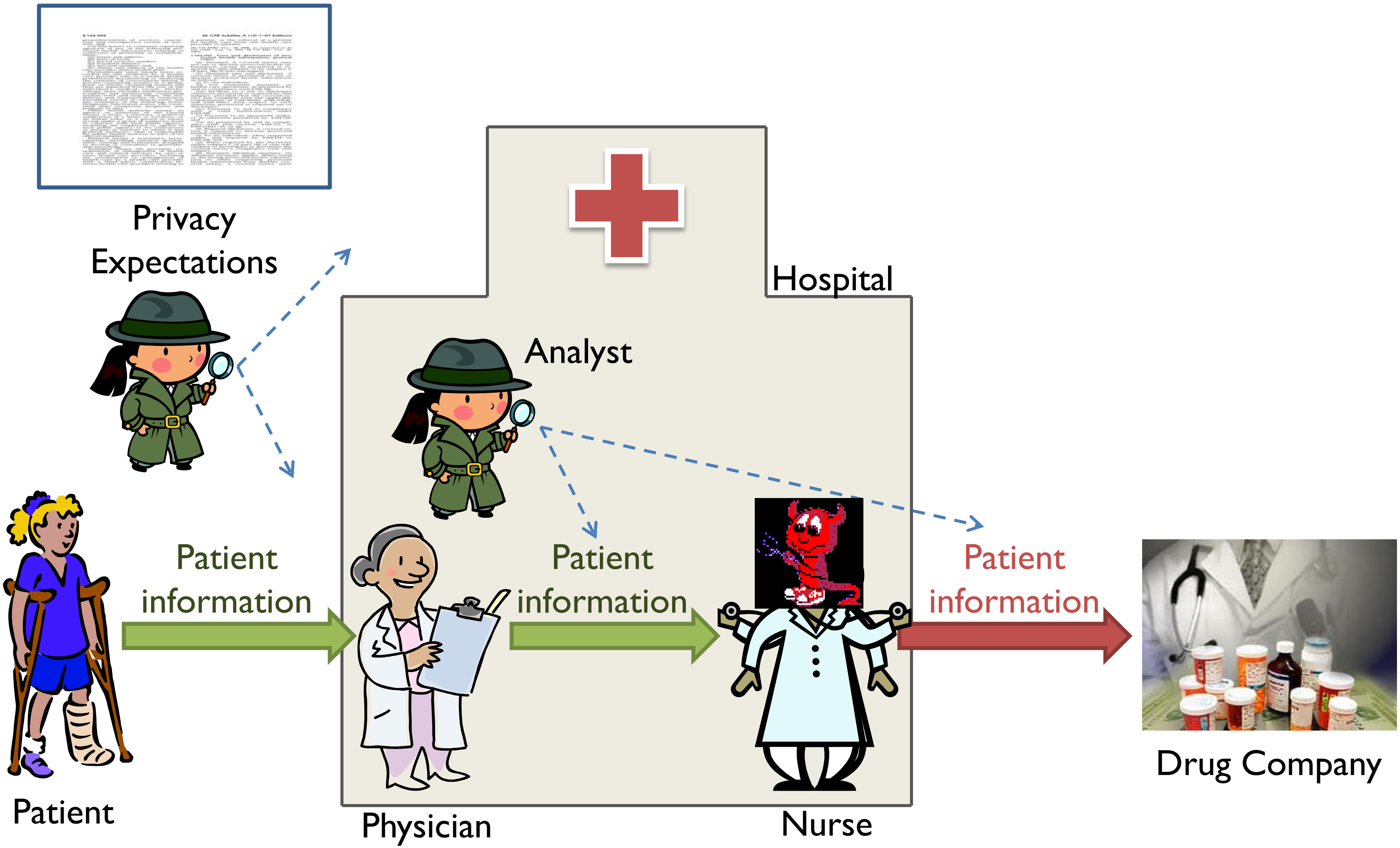
Privacy through Accountability

Anupam Datta

CMU

Fall 2013

Healthcare Privacy



Auditing

- Permissive real time access policy
- Inspect accesses after occurrence
- Find and punish policy violators

- Combining automated and human audits
  - Example: FairWarning Audit Tool flags all celebrity record accesses as suspicious

Automated Audit of Purpose Restrictions

With M. C. Tschantz (CMU → Berkeley) and
J. M. Wing (CMU → MSR)
2012 IEEE Symposium on Security & Privacy

Goal

- Give a semantics to
  - **“Not for” purpose restrictions**
  - **“Only for” purpose restrictions**

  that is parametric in the purpose

- Provide automated enforcement of purpose restrictions for that semantics

Purpose Restrictions in Privacy Policies

Not for

- Yahoo!'s practice is **not** to use the content of messages […] **for** marketing **purposes**.

Only for

- By providing your personal information, you give [Social Security Administration] consent to use the information **only for** the **purpose** for which it was collected.

Purpose Restrictions are Ubiquitous

- OECD's Privacy Guidelines
- US Privacy Laws
  – HIPAA, GLBA, FERPA, COPPA,…
- EU Privacy Directive
- Organizational Privacy Policies
  – Google, Facebook, Yahoo,…
  – Hospitals, banks, educational institutions, govt
  – Defense: Mission-based information access

X-ray taken

Send record

No diagnosis
by drug company

Add x-ray

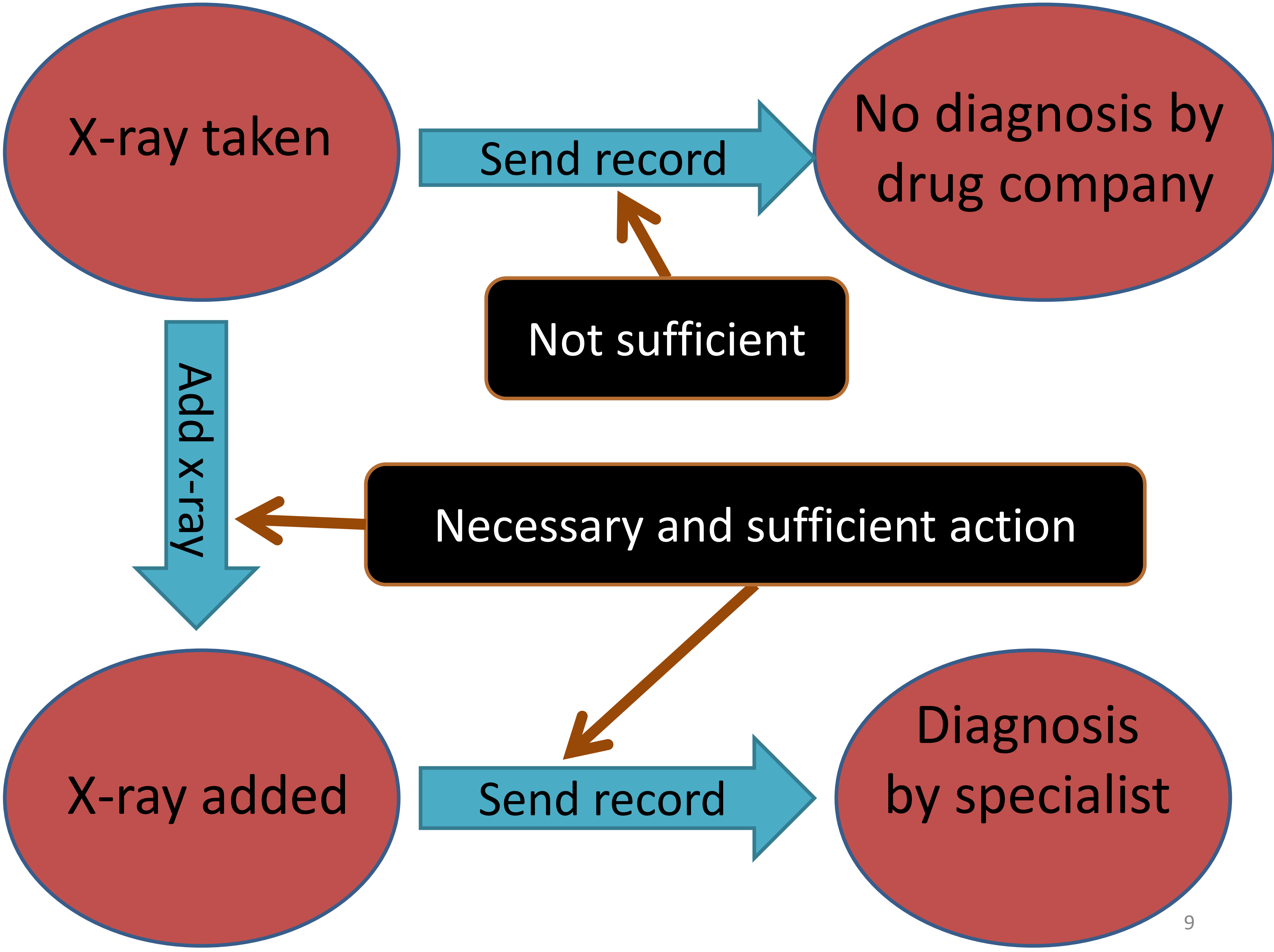
Medical
Record

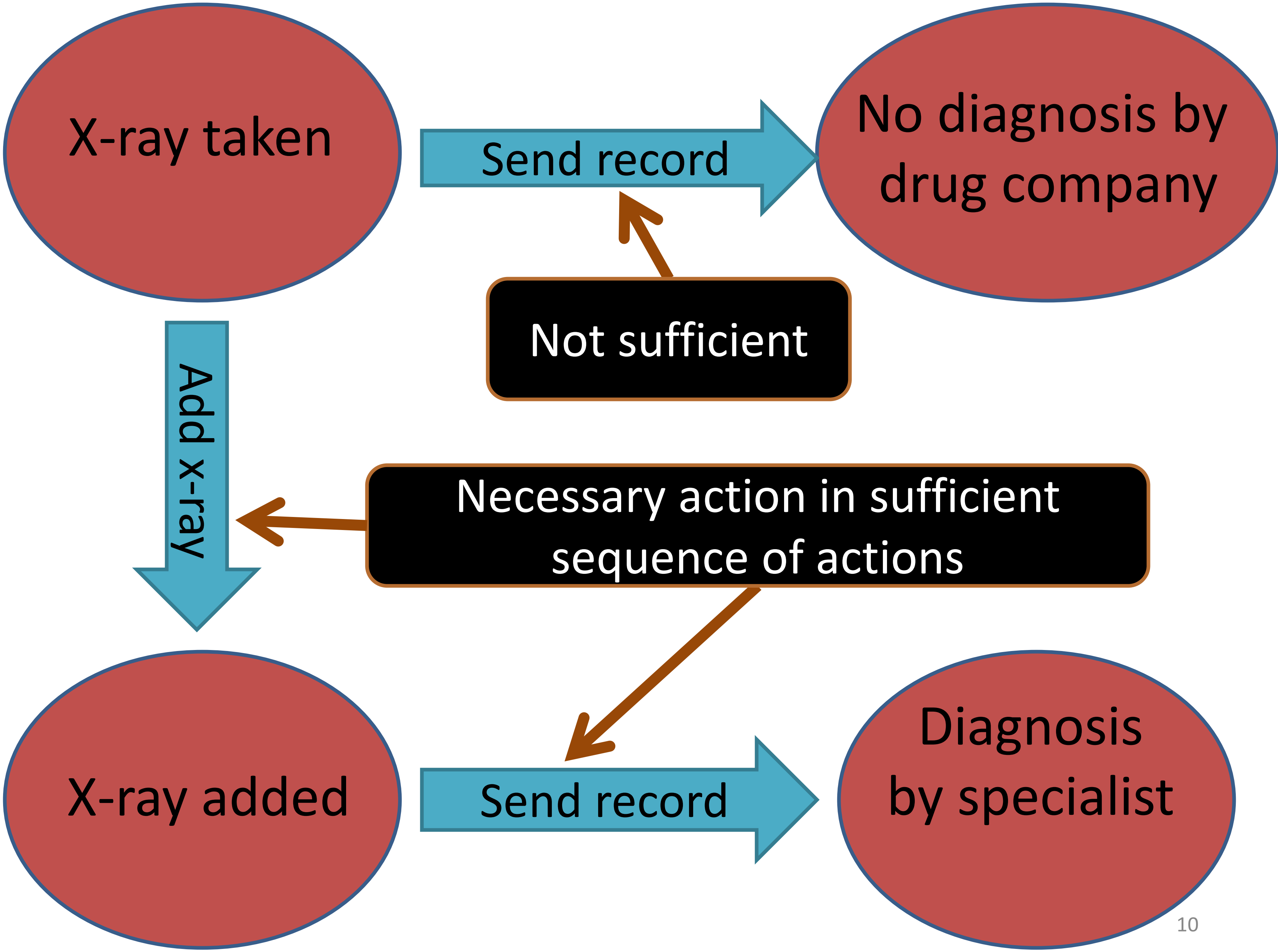
Med records
used only for
diagnosis

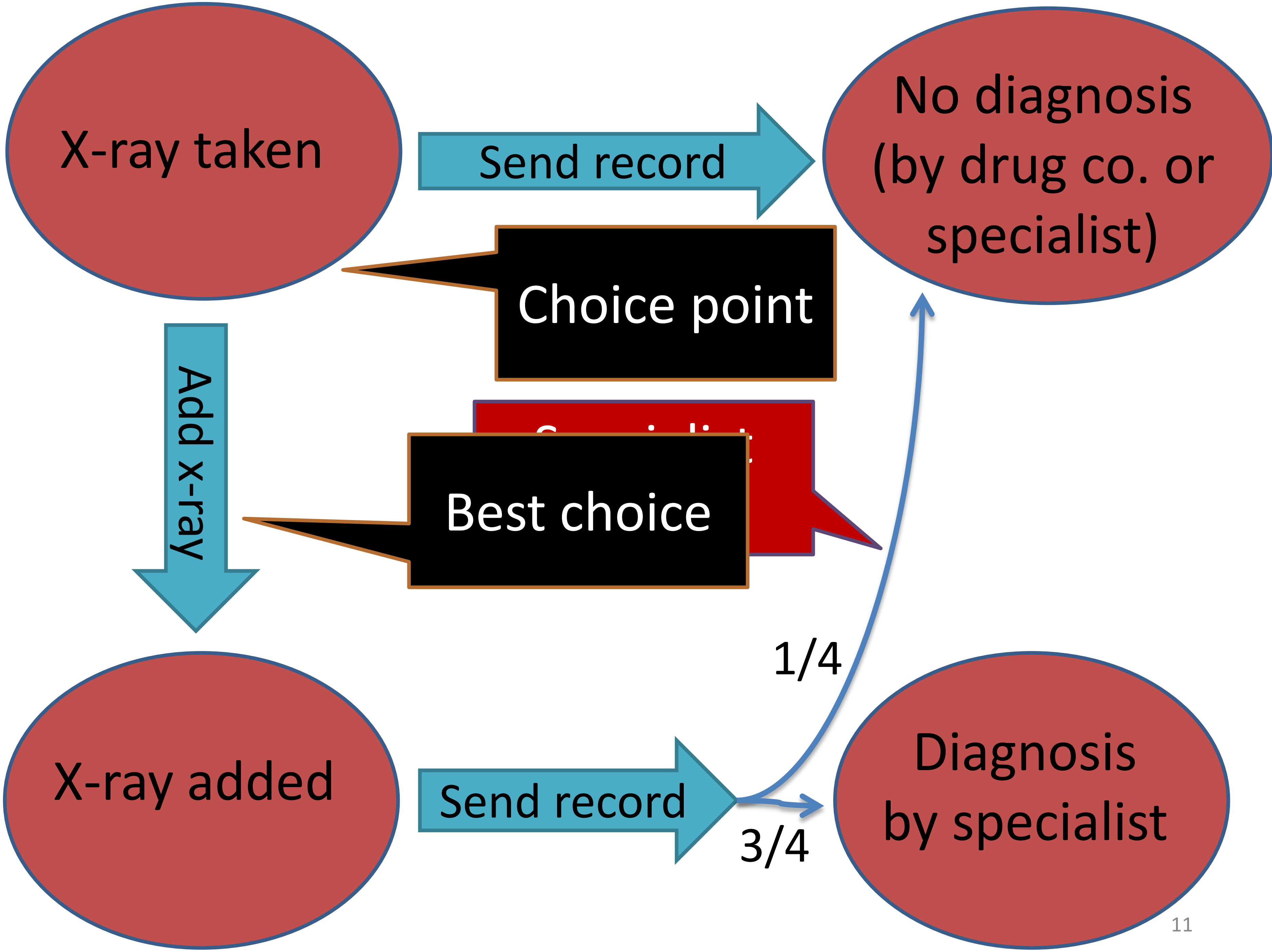
X-ray added

Send record

Diagnosis
by specialist


X-ray taken
Send record
No diagnosis by drug company
Not sufficient
Add x-ray
Necessary and sufficient action
X-ray added
Send record
Diagnosis by specialist

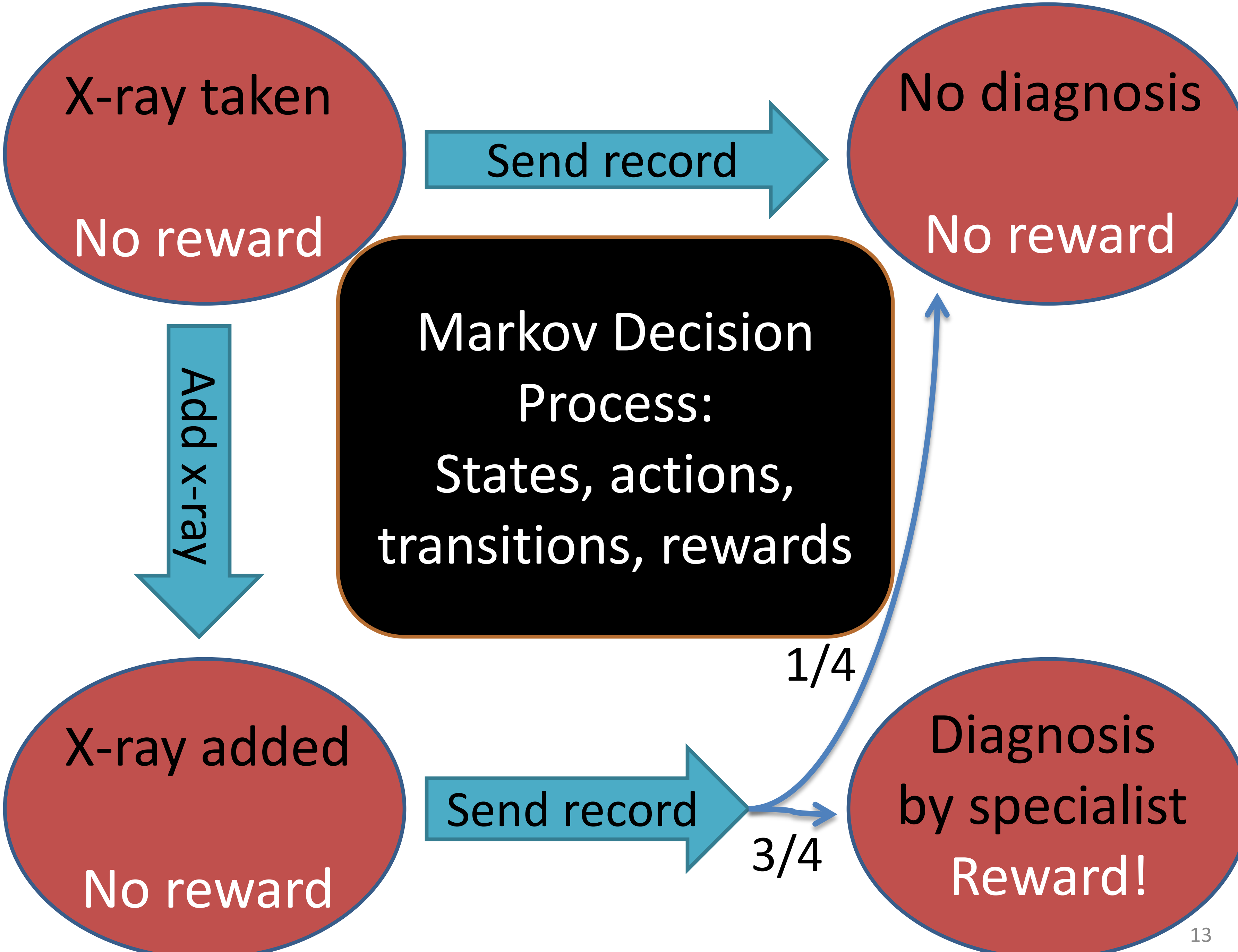

X-ray taken
Send record
No diagnosis by drug company
Not sufficient
Add x-ray
Necessary action in sufficient sequence of actions
X-ray added
Send record
Diagnosis by specialist


X-ray taken
Send record
No diagnosis (by drug co. or specialist)
Choice point
Add x-ray
Best choice
1/4
X-ray added
Send record
3/4
Diagnosis by specialist

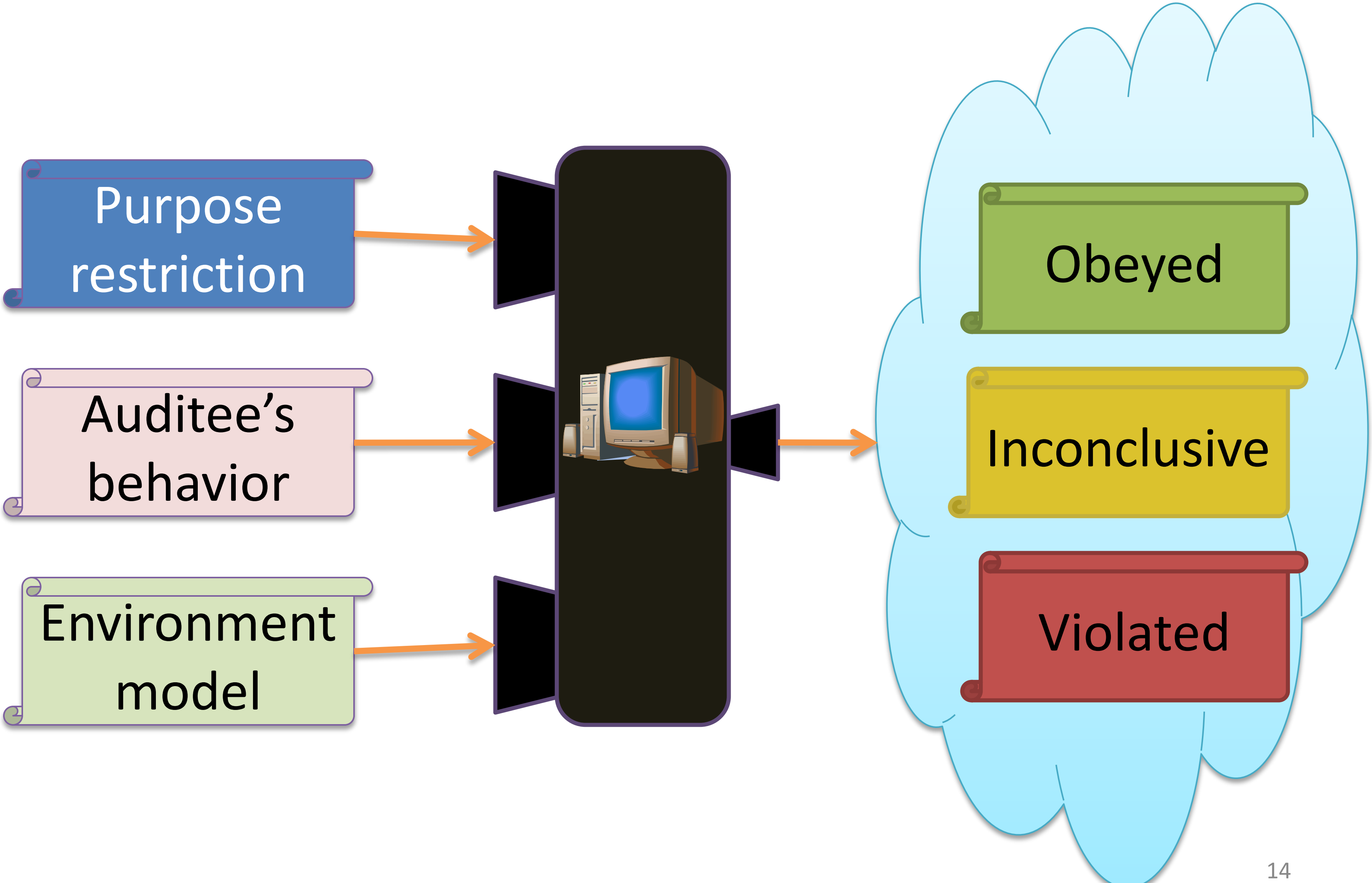
Planning

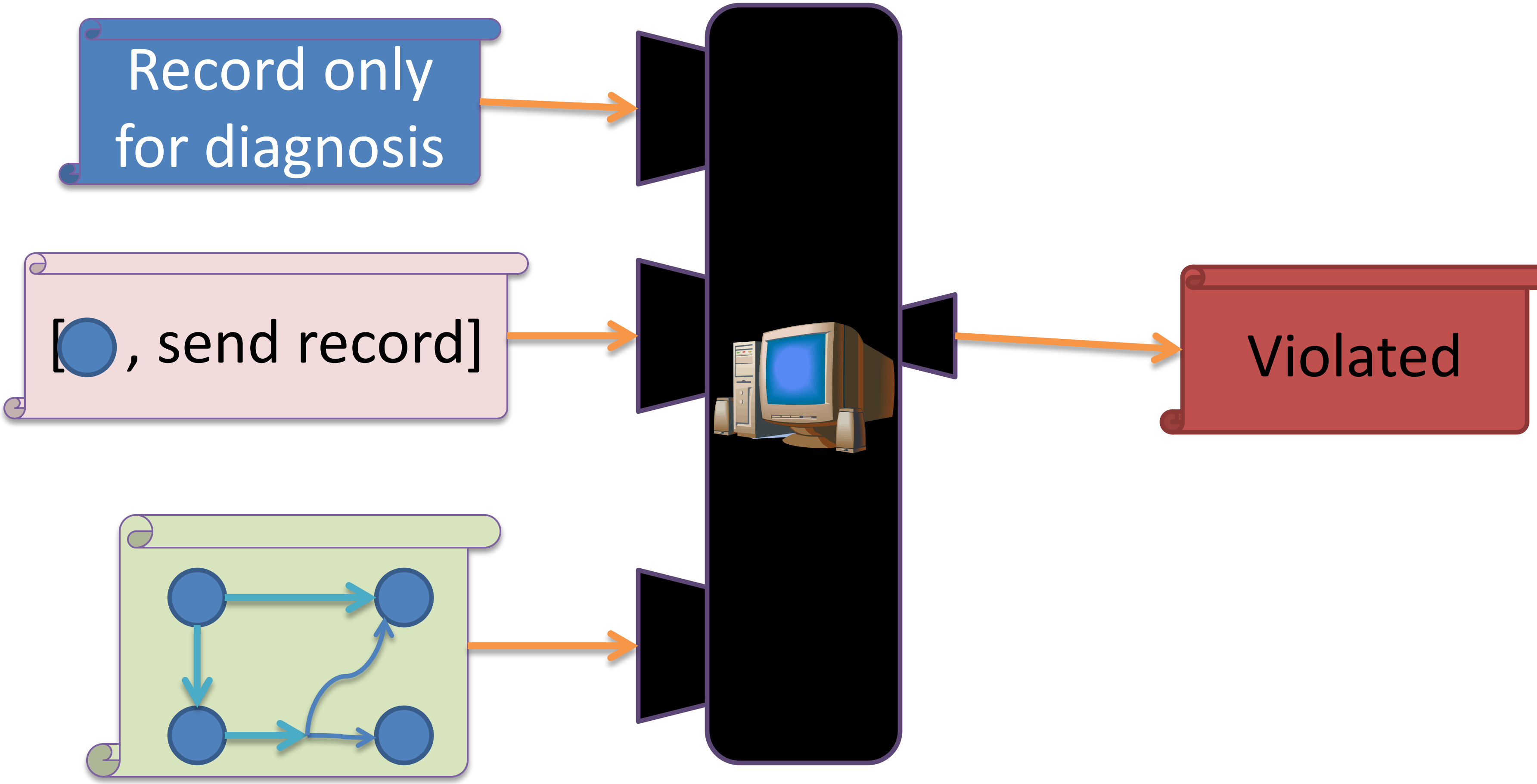
Thesis: An action is for a purpose iff that action is part of a plan for furthering the purpose

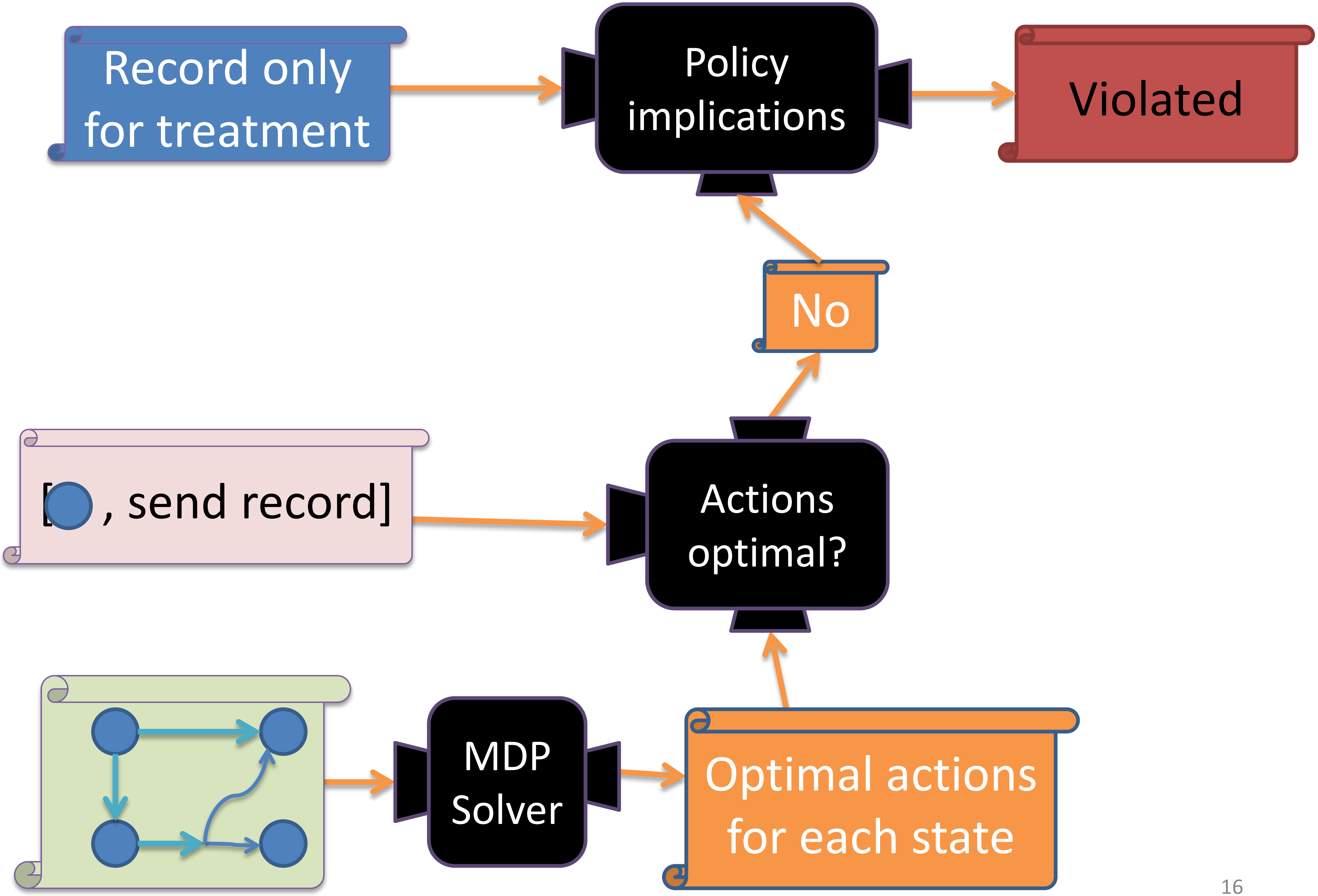
i.e., always makes the best choice for furthering the purpose


X-ray taken
No reward
Send record
No diagnosis
No reward
Markov Decision Process:
States, actions, transitions, rewards
Add x-ray
1/4
X-ray added
No reward
Send record
3/4
Diagnosis by specialist
Reward!

Auditing


Purpose restriction
Auditee's behavior
Environment model
Obeyed
Inconclusive
Violated

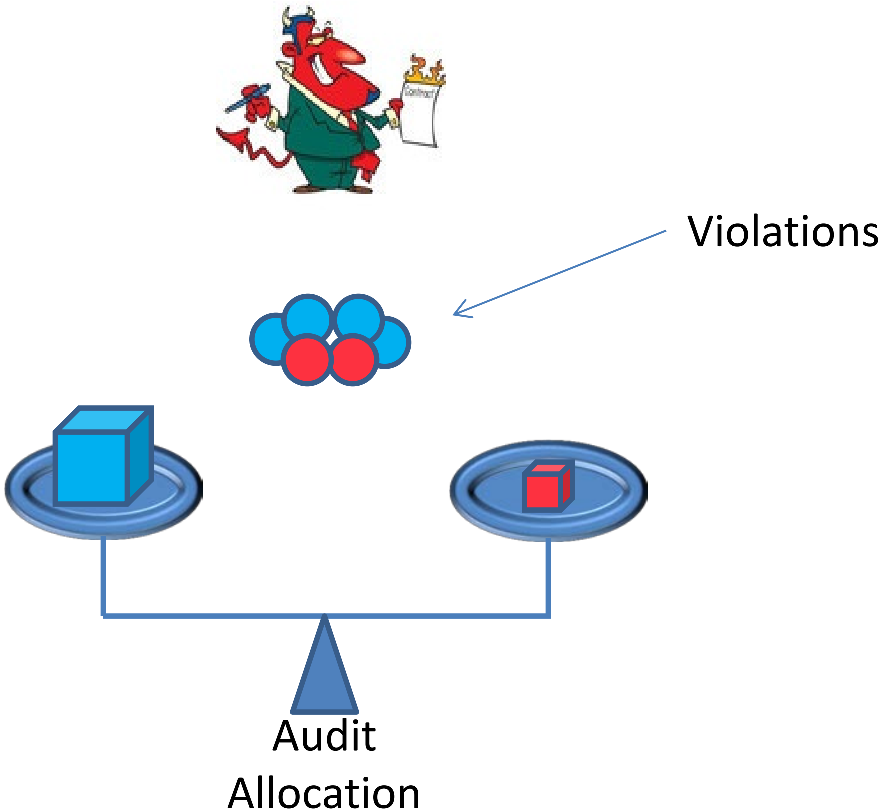

Record only
for diagnosis
[, send record]
Violated



Record only for treatment
Policy implications
Violated
No
[, send record]
Actions optimal?
MDP Solver
Optimal actions for each state

Audit Games:
Resource Allocation for Human Auditors

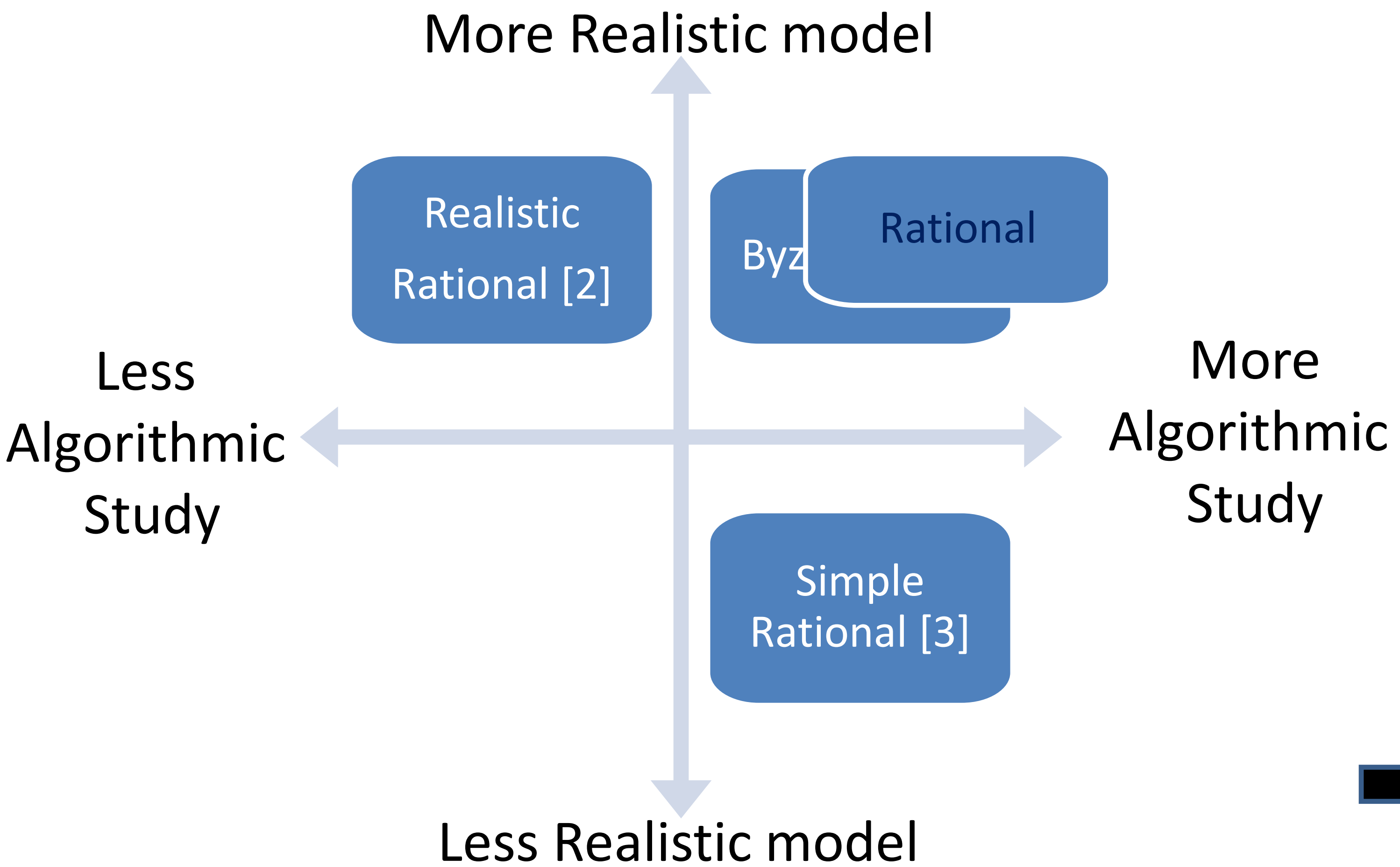
Game Interaction



Optimal allocation depends on adversary behavior

Game model appropriate for Auditing

Problem Space



[1]J. Blocki, N. Christin, A. Datta, A. Sinha, Regret Minimizing Audits, Computer Security Foundations , June11
[2]J. Blocki, N. Christin, A. Datta, A. Sinha, Audit Mechanisms for Prov. Risk Mngmt. & Accountable Data Gov., GameSec Nov12
[3]J. Blocki, N. Christin, A. Datta, A. Procaccia, A. Sinha, Audit Games, Int. Joint Conf. on Artificial Intelligence, Aug13

Outline of the talk

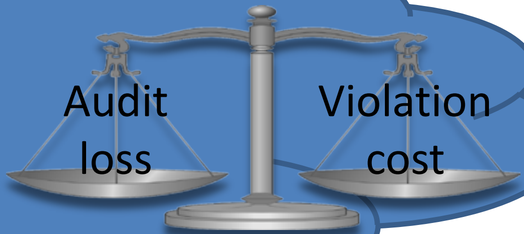
- Completed work
  - Byzantine adversary model [1]
  - Simple rational adversary model [3]
- Future Work
  - Extending the simple rational adversary

[1]J. Blocki, N. Christin, A. Datta, A. Sinha, Regret Minimizing Audits, Computer Security Foundations , June11
[3]J. Blocki, N. Christin, A. Datta, A. Procaccia, A. Sinha, Audit Games, Int. Joint Conf. on Artificial Intelligence, Aug13

Model/Algorithm by Example

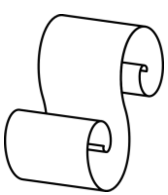
Auditor

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



Access divided into 2 types

Loss from 1 violation (internal, external)

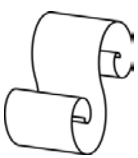


100 accesses

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	0	10	20	30
	30	20	10	0
Weights	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 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 observed and estimated loss

Byzantine model

- k types of target
  - $\vec{n} = n_1, \ldots, n_k$ targets
  - $\vec{s}$ inspections, $\vec{v}$ violations
  - $\overrightarrow{\boldsymbol{O}}$ violations – parameterized by $\vec{n}, \vec{s}, \vec{v}$
  - Fixed probability p of external detection
- Defender action - Inspections: $\vec{s}$ chosen at random
- Adversary action - Violations: $\vec{v}, \vec{n}$
- Repeated game
  - Rounds correspond to audit cycle

Utilities

- $\boldsymbol{U}(\vec{s}, \overrightarrow{\boldsymbol{O}}) = \underbrace{\sum_k U_1(s_k)}_{\text{Audit Cost}} + \underbrace{\sum_k U_2(\boldsymbol{O}_k)}_{\text{Violation Cost}}$

- Average utility over T rounds $= \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{U}(\vec{s}^{t}, \overrightarrow{\boldsymbol{O}}^{t})$

- Adversary utility unknown

Regret by Example

		Org 1	Org 2
Emp	3,1	\$5	\$6
	3, 2	\$0	\$5

Strategy: outputs an action for every round

$$Total\ Regret(s, s_1) = -5\ - (-6) = 1$$

$$regret(s, s_1) = \frac{1}{2}$$

Players	Round 1	Round 2	Total Payoff
• Emp • Org: s	• 3,1 • 2 (\$6)	• 3, 2 • 1 (\$0)	• Unknown • \$6
Org : s_1	1 (\$5)	1 (\$0)	\$5

Meaning of Regret

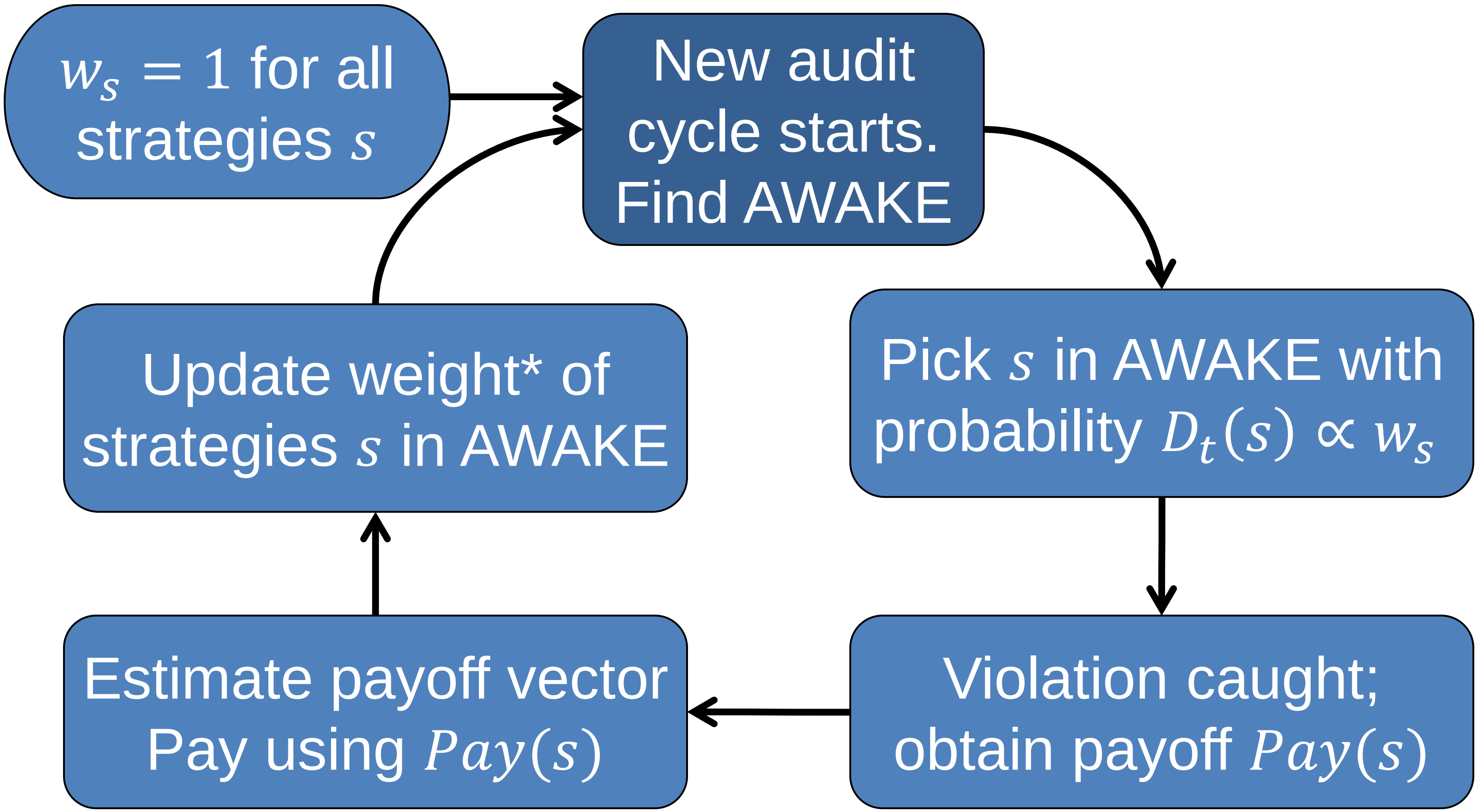
- Low regret of s w.r.t. s_1 means s performs as well as s_1

- Desirable property of an audit mechanism
  - Low regret w.r.t. a set of strategies S
  - $\max_{s' \in S} regret(s, s') \to 0 \; as \; T \to \infty$

Known Algorithms

- MWU is a standard algorithm with regret bound
  - $2\sqrt{\frac{\log(N)}{T}}$
  - 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 (sleeping experts)

Regret Minimizing Algorithm



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

Guarantees of RMA

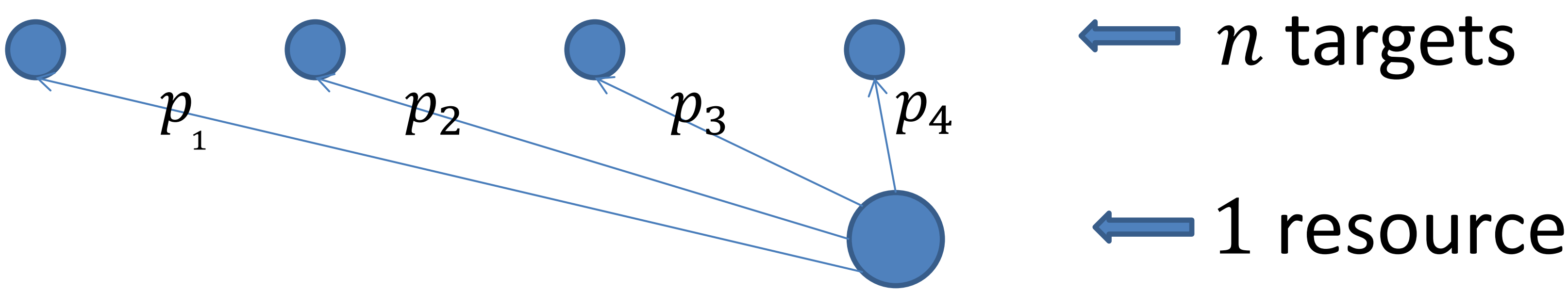
- With probability $1 - \epsilon$ RMA achieves the regret bound

$$2\sqrt{\frac{2\log(N)}{T}} + \frac{2\log(N)}{T} + 2\sqrt{\frac{2\log(4N/\epsilon)}{T}}$$

  - N is the set of strategies
  - T is the number of rounds
  - All payoffs scaled to lie in [0,1]

- Better bound than existing algorithm (under mild assumptions)

Simple Rational Model



- Adversary commits one violation
- If a violation is detected, adversary is fined $\$x$
- Utility when target t_i is attacked
  - $p_i\, U_{a,D}(t_i) + (1 - p_i)U_{u,D}(t_i) - a_0 x$
  - $p_i\, (U_{a,A}(t_i) - x) + (1 - p_i)U_{u,A}(t_i)$

Utility when audited   Utility when unaudited

Stackelberg Equilibrium Concept

- Defender commits to a randomized resource allocation strategy (p_i's and x)
- Adversary plays best response to that strategy
- For defender Stackelberg better than Nash eq.
- Goal
  - Compute optimal defender strategy

Computing Optimal Defender Strategy

Solve optimization problems P_i for all $i \in \{1, \ldots, n\}$ and pick the best solution

$$\max p_i\, U_{a,D}(t_i) + (1 - p_i) U_{u,D}(t_i) - a_0 x$$

subject to

$\forall j \in \{1, \ldots, n\}$

$$p_j\,(U_{a,A}(t_j) - x) + (1 - p_j) U_{u,A}(t_j) \leq p_i\left(U_{a,A}(t_i) - x\right) + (1 - p_i) U_{u,A}(t_i)$$

and p_i's lie on the probability simplex

and $0 \leq x \leq 1$

Quadratic
Non-convex

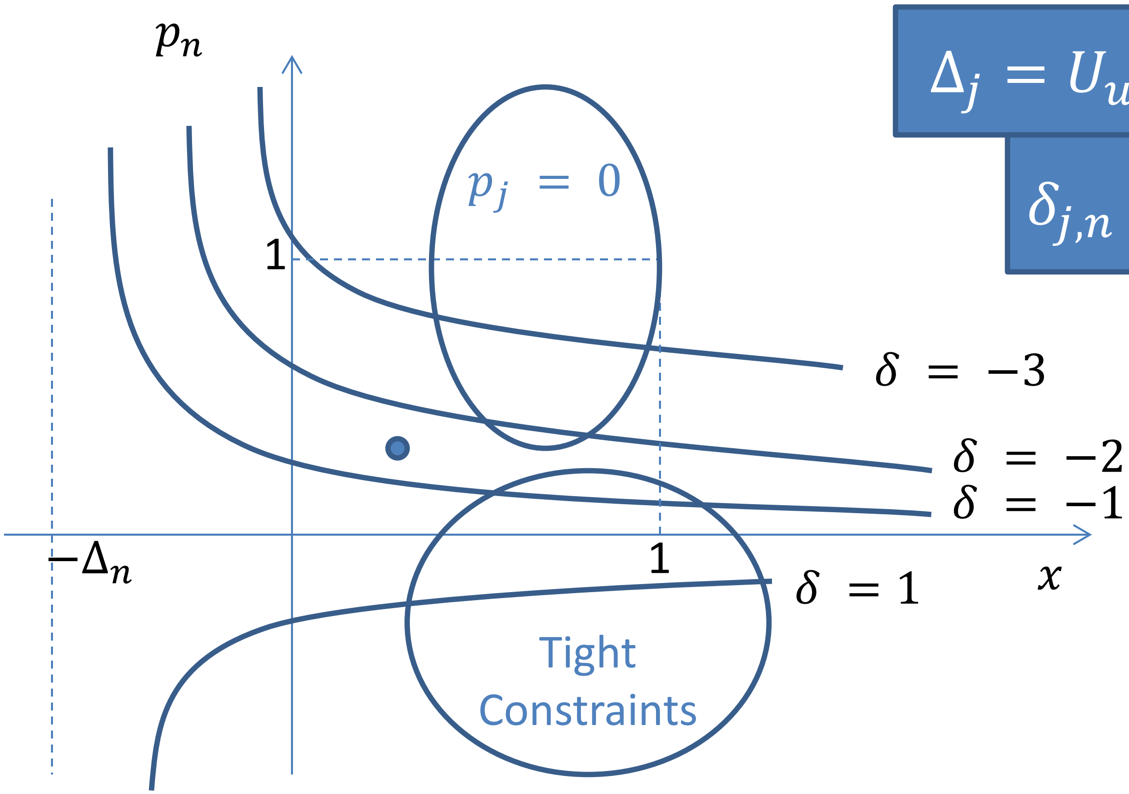
Properties of Optimal Point

- Rewriting quadratic constraints

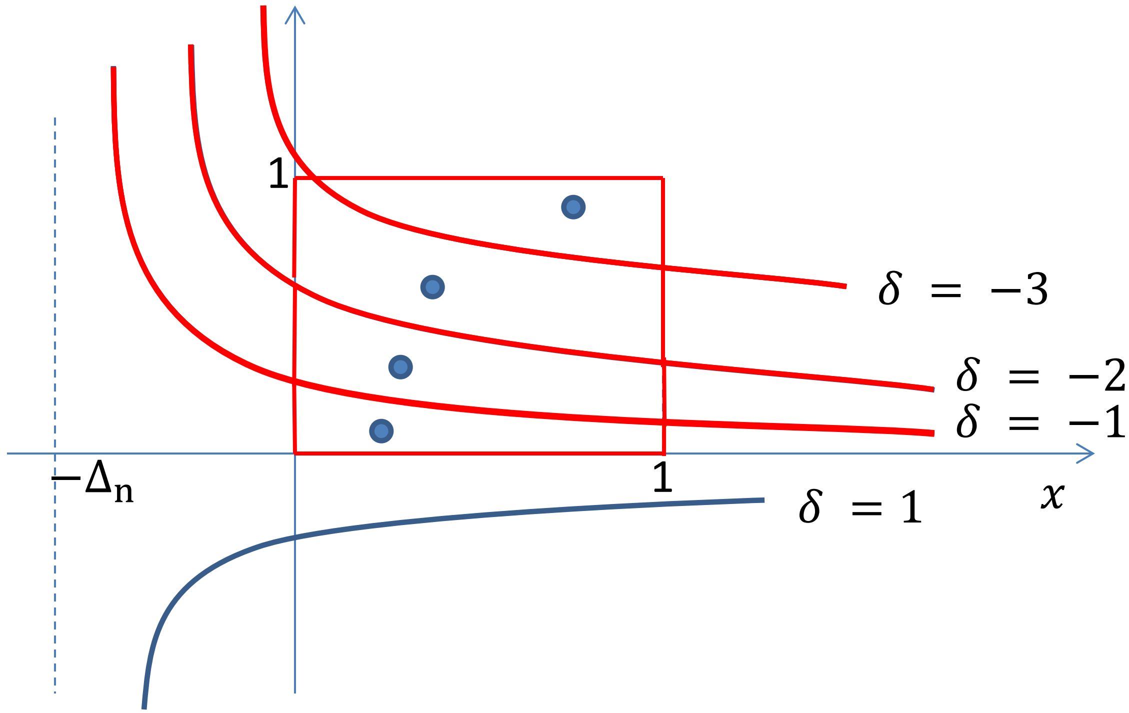
$$p_j(-x - \Delta_j) + p_n(x + \Delta_n) + \delta_{j,n} \leq 0$$

$$\Delta_j = U_{u,A}(t_j) - U_{a,A}(t_j) \geq 0$$

$$\delta_{j,n} = U_{u,A}(t_j) - U_{u,A}(t_n)$$



Main Idea in Algorithm



- Iterate over regions, solve sub-problems EQ_j
  - Set probabilities to zero for curves that lie above & make other constraints tight
- Pick best solution of all EQ_j

Solving Sub-problem EQ_j

1. $p_j(-x - \Delta_j) + p_n(x + \Delta_n) + \delta_{j,n} = 0$
   - Eliminate p_j to get a equation in p_n and x only
2. Express p_n as a function $f(x)$
   - Objective becomes a polynomial function of x only
3. Find x where derivative of objective is zero & constraints are satisfied
   - Local maxima
4. Find x values on the boundary
   - Found by finding intersection of p_n = f(x) with the boundaries
   - Other potential points of maxima
5. Take the maximum over all x values from steps 3,4

Main Theorem

- *The problem can be approximated to an additive* ϵ *factor in time* $O\left(n^5 K + n^4 \log\left(\frac{1}{\epsilon}\right)\right)$ *using only the splitting circle method, where K is the bit precision of inputs.*

Background: Security Games

- Game model for physical sec. - extensively studied
  - LAX airport deployment
  - Air marshals deployment
- High level (basic) model
  - n targets defended by m resources
  - Schedules: constraints on use of resources
    - given as function from resources to sets of targets
  - Stackelberg equilibrium
  - No punishments

Extending the simple rational model

- More than one defender resources
  - Schedules
- More than one target attacked by adversary
  - Simple case: less than a constant number of attacks
- Cost of resources/Budget
  - Resources are not given, but, cost money to buy

Conclusion

A resource-constrained auditor's interaction with an adaptive adversary can be formalized using game-theoretic models and audit algorithms can be designed that provably optimize the defender's utility function in these models against Byzantine and rational adversaries

- Questions?