A Guided Tour Through

Adversarial Risk Analysis

Concepts, Applications and Challenges

David Ríos Insua

Royal Academy of Sciences

UC3M, May'13

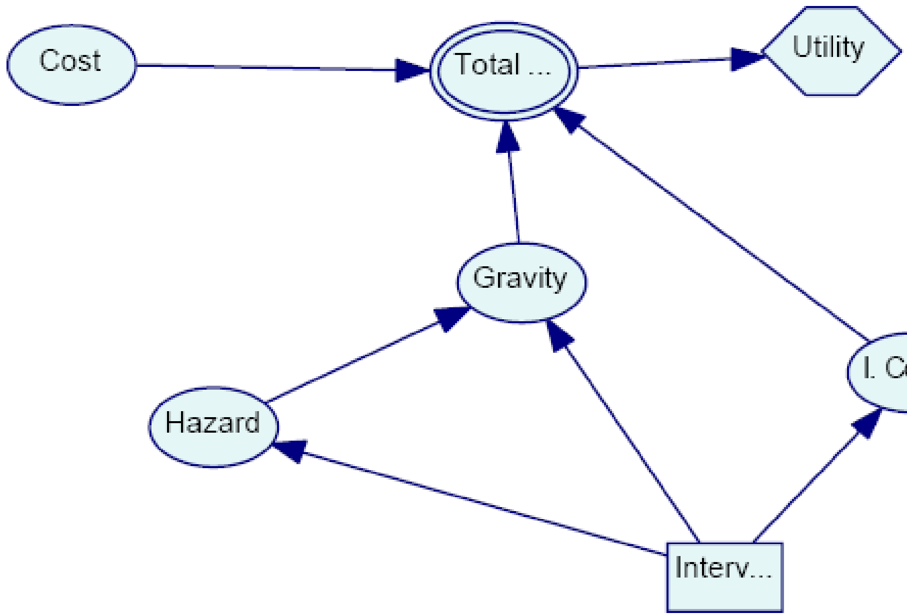
D. Banks (Duke), J. Rios (IBM), J.Cano, J. Williams, A. Schmitz (SECONOMICS), P. Garcia, A. Redondo (URJC), D. García (Aisoy)

Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- **Security**
- Intelligent interfaces
- Challenges

Risk management

Intervention to be chosen:



Interventions tend to reduce the likelihood of hazard appearance and its gravity... but they also entail a cost

Gain through managed risk

Choose the intervention which provides the biggest gain, if it is sufficiently big...

Which is the best security resource allocation in a city?

City as a map with cells

Each cell has a value

For each cell, a predictive model of delictive acts

Allocate security resources (constraints)

For each cell predict the impact of resource allocation

Optimal resource allocation

NB: The bad guys also operate intelligent and organisedly!!!

SECONOMICS (Metro Barcelona, UK Grid, Anadolu Airport)

Which is the best HW/SW maintenance for the university ERP?

Model HW/SW system (interacting HW and SW blocks)

Forecast block reliability

Forecast system reliability

Design maintenance policies

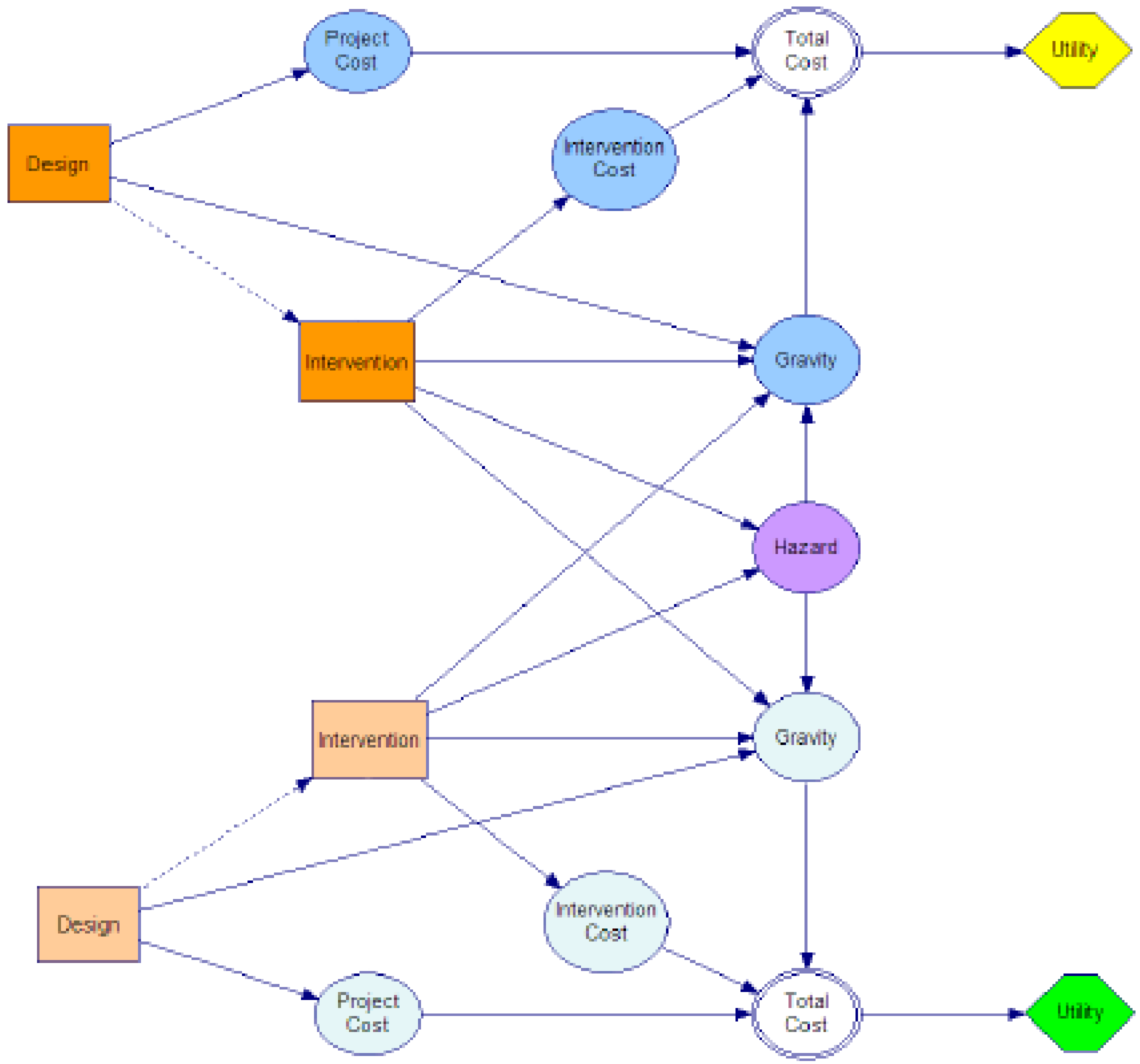
Forecast impact on reliability (and costs)

Optimal maintenance policy

NB: Again, what happens with the bad guys attacking our system?

RIESGOS (MICINN), RIESGOS-CM (CM)

Adversarial risk analysis



Outline

- From risk analysis to adversarial risk analysis
- **Motivation**
- Sequential games
- Simultaneous games
- Auctions
- Security
- Intelligent interfaces
- Challenges

Motivation

- Tradiitonal RA extended to include adversaries ready to increase our risks

- S-11, M-11 lead to large security investments globally, some of them criticised
- Many modelling efforts to efficiently allocate such resources
- Parnell et al (2008) NAS review
  - Standard reliability/risk approaches not take into acocunt intentionality
  - Game theoretic approaches. Common knowledge assumption…
  - Decision analytic approaches. Forecasting the adversary action…
- Merrick, Parnell (2011) review approaches commenting favourably on Adversarial Risk Analysis

Adversarial Risk Analysis

- A framework to manage risks from actions of intelligent adversaries (DRI, Rios, Banks, JASA 2009)

- One-sided prescriptive support
  - Use a SEU model
  - Treat the adversary's decision as uncertainties

- Method to predict adversary's actions
  - We assume the adversary is a *expected utility maximizer*
    - Model his decision problem
    - Assess his probabilities and utilities
    - Find his action of maximum expected utility
  - But other *descriptive* models are possible

- Uncertainty in the Attacker's decision stems from
  - *our* uncertainty about his probabilities and utilities
  - but this leads to a hierarchy of nested decision problems

(random, noninformative, k-level, heuristic, mirroring argument) vs (common knowledge)

Adversarial Risk Analysis

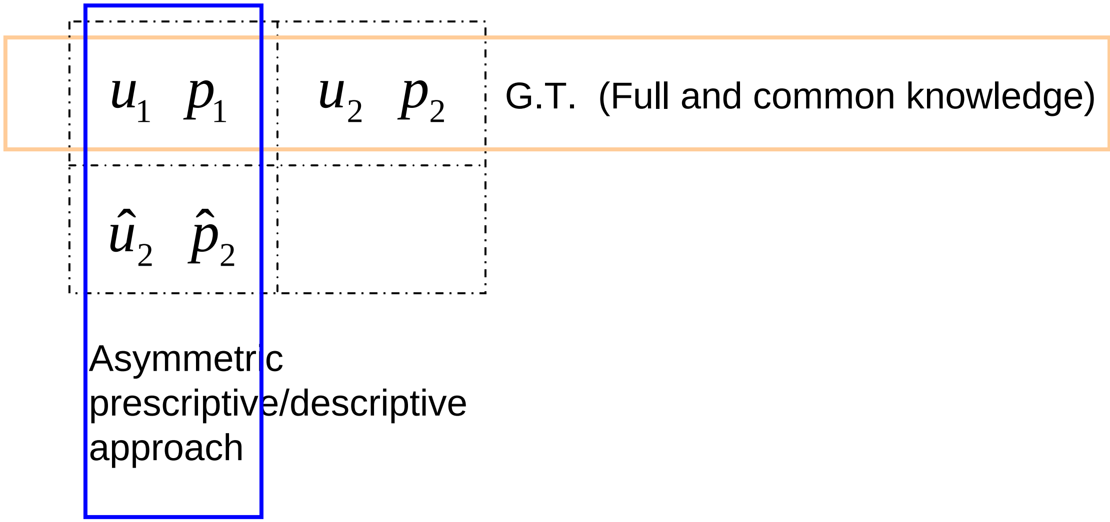
- ARA applications to counterterrorism models (Rios, DRI, 2009, 2012 Risk Analysis)
  - Sequential Defend-Attack
  - Simultaneous Defend-Attack
  - Sequential Defend-Attack-Defend
  - Sequential Defend-Attack with private information

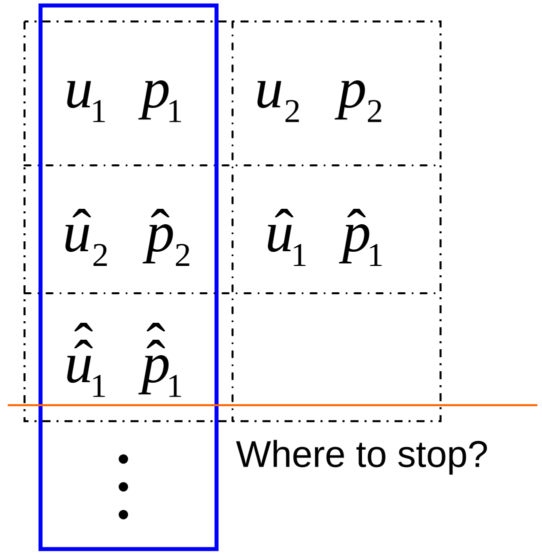
- Somali pirates case (Sevillano, Rios, DRI, 2012 Decision Analysis)
- Routing games (anti IED war) (Wang, Banks, 2011)

- Borel games (Banks, Petralia, Wang, 2011)
- Auctions (DRI, Rios, Banks, 2009; Rothkopf, 2007)

- Kadane, Larkey (1982), Raiffa (1982), Lippman, McCardle (2012)
- Stahl and Wilson (1994,1995) D. Wolpert (2012)
- Rotschild, MacLay, Guikema (2012)

Adversarial risk analysis

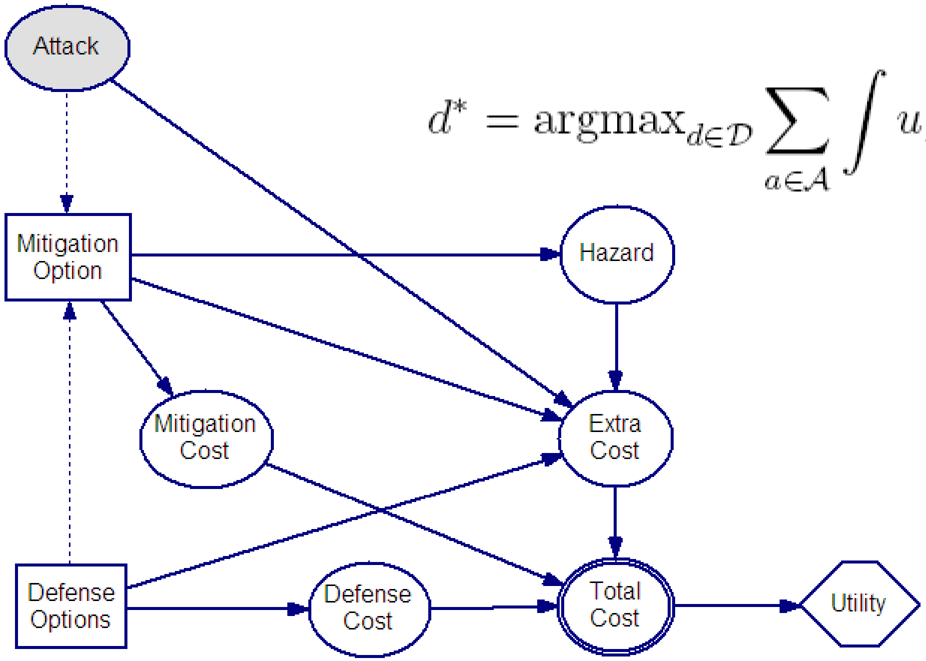




Asymmetric prescriptive/descriptive approach

- Bayesian approach (Raiffa, Kadane, Larkey…)
  - Prescriptive advice to one party conditional on a (probabilistic) description of how others will behave
  - Treat the other participant's decisions as uncertain

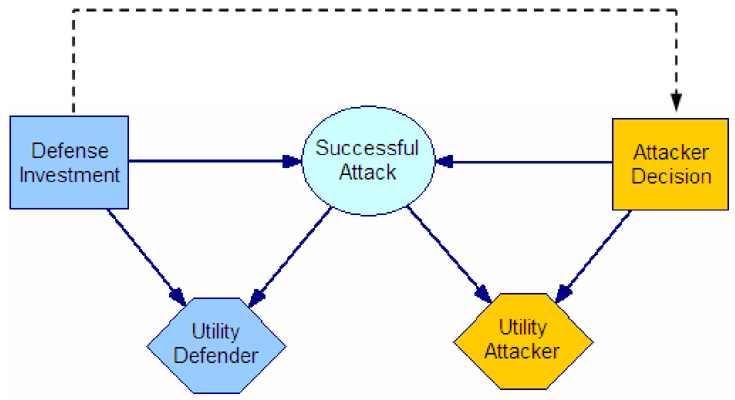
$$d^* = \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \int u_D(c) \pi_D(c|a, d) p_D(a|d) dc$$

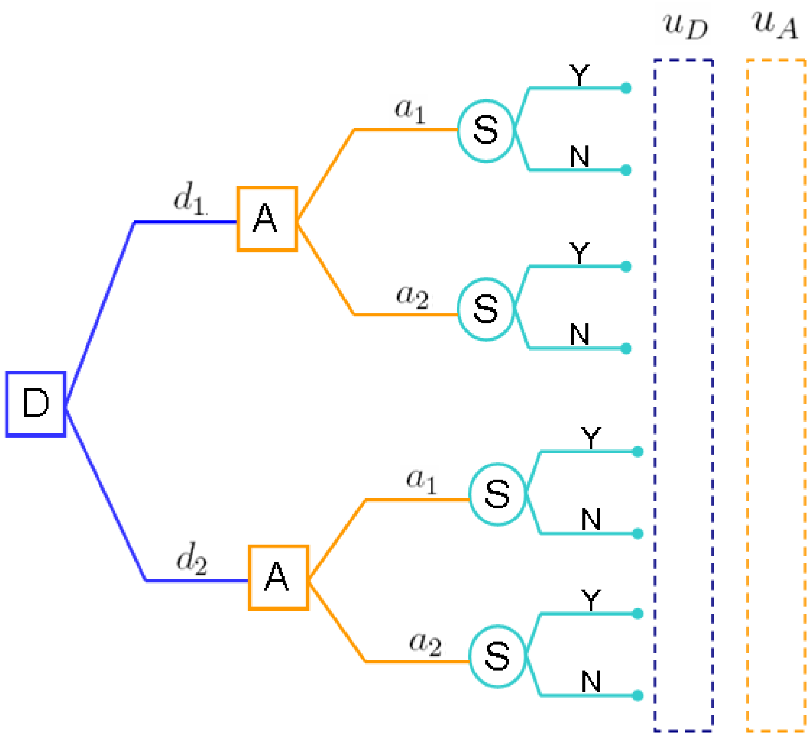


Outline

- From risk analysis to adversarial risk analysis
- Motivation
- **Sequential games**
- Simultaneous games
- Auctions
- Security
- Intelligent interfaces
- Challenges

Sequential games: First Defender, afterwards Attacker





$$a^*(d) = \arg\max_{a \in X_A} u_A(d, a)$$

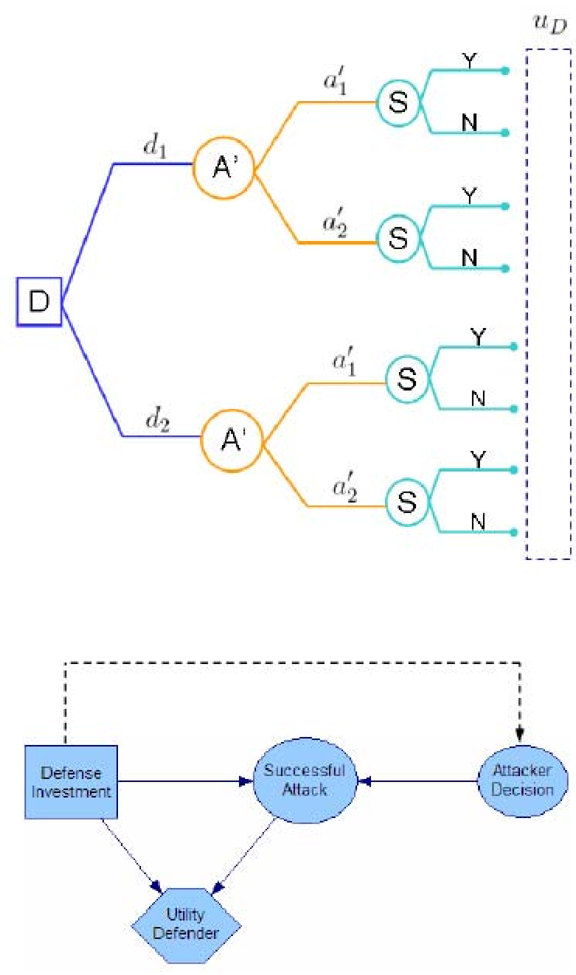
$$d^* = \arg\max_{d \in X_D} u_A(d, a^*(d))$$

Nash Solution: $(d^*, a^*(d^*))$

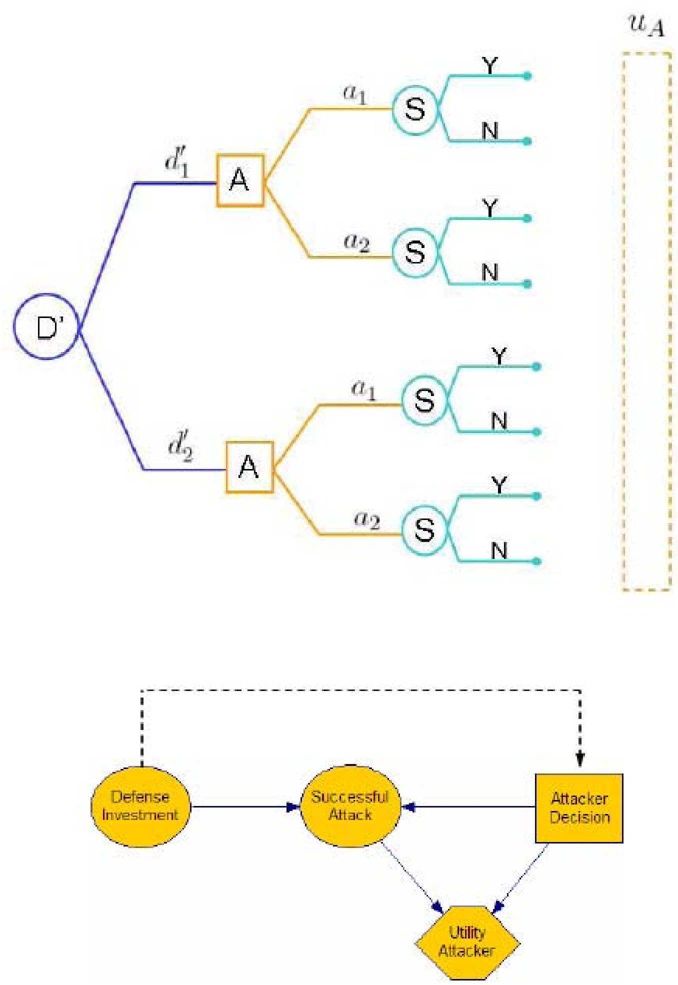
Standard
Game Theory Analysis

The sequential game: Supporting the Defender

Defender problem

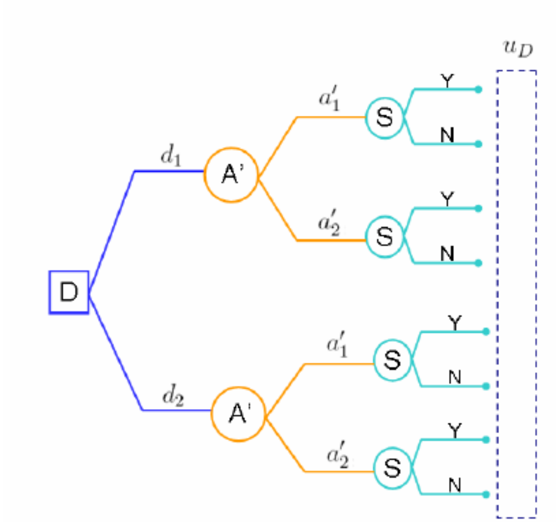


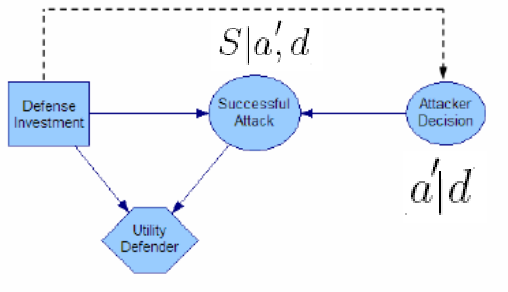
Defender's view of Attacker problem



Supporting the Defender

Defender problem	Defender's solution





$$\psi_D(d,a') = u_D(d, S=Y)\, p_D(S=Y \mid X_D = d, X'_A = a') + u_D(d, S=N)\, p_D(S=N \mid X_D = d, X'_A = a')$$

$$\psi_D(d) = \psi_D(d, a'_1)\, p_D(a'_1 \mid d) \;+\; \psi_D(d, a'_2)\, p_D(a'_2 \mid d)$$

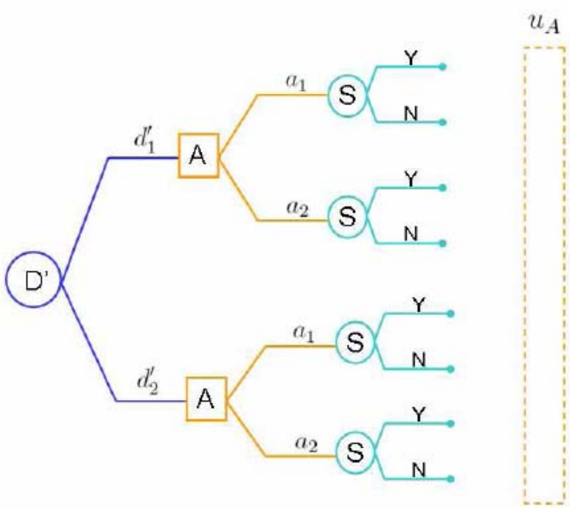
$$d^* = \arg\max_{d \in X_D} \psi_D(d)$$

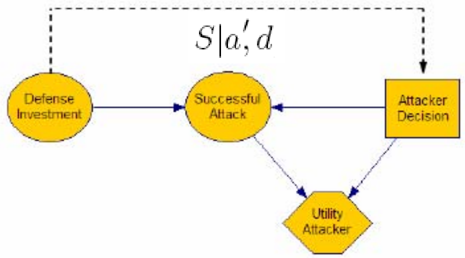
Modeling input: $p_D(S|a',d)$ $p_D(a'|d)$??

Supporting the Defender: The assessment problem

Defender's view of Attacker problem	Elicitation of $p_D(a'|d)$

Defender's view of Attacker problem





Elicitation of $p_D(a'|d)$

A is a EU maximizer

D's beliefs about $(\widehat{u}_A, \widehat{p}_A) \sim F$

$$\hat{\psi}_A(d', a) = \hat{u}_A(a, S = Y)\, \hat{p}_A(S = Y \mid X'_D = d',\ X_A = a) + \hat{u}_A(a, S = N)\, \hat{p}_A(S = N \mid X'_D = d',\ X_A = a)$$

$$\hat{\psi}_A \sim \widehat{\Psi}_A$$

$$p_D(a'|d) = Pr\left[a' = \arg\max_{x \in X'_A} \widehat{\Psi}_A(d, x)\right]$$

<u>MC simulation</u>

$$\hat{p}_D(a|d) \approx n^{-1} \sum_i \#\{a = \text{argmax}_{x \in \mathcal{A}} \hat{\psi}^i_A(x, d)\}$$

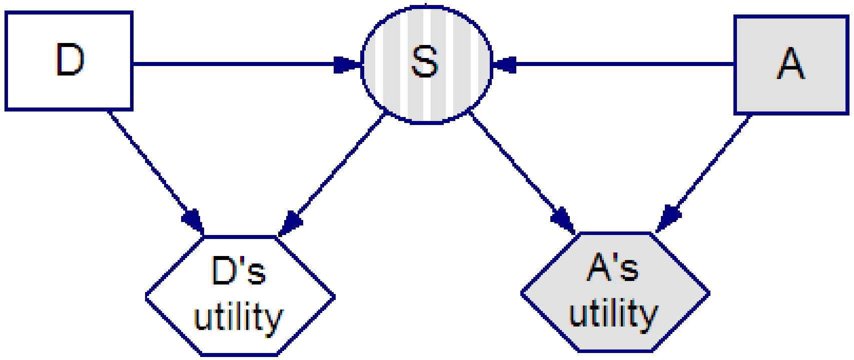
where $\hat{\psi}^i_A \sim \widehat{\Psi}_A,\ i = 1, \ldots, n$

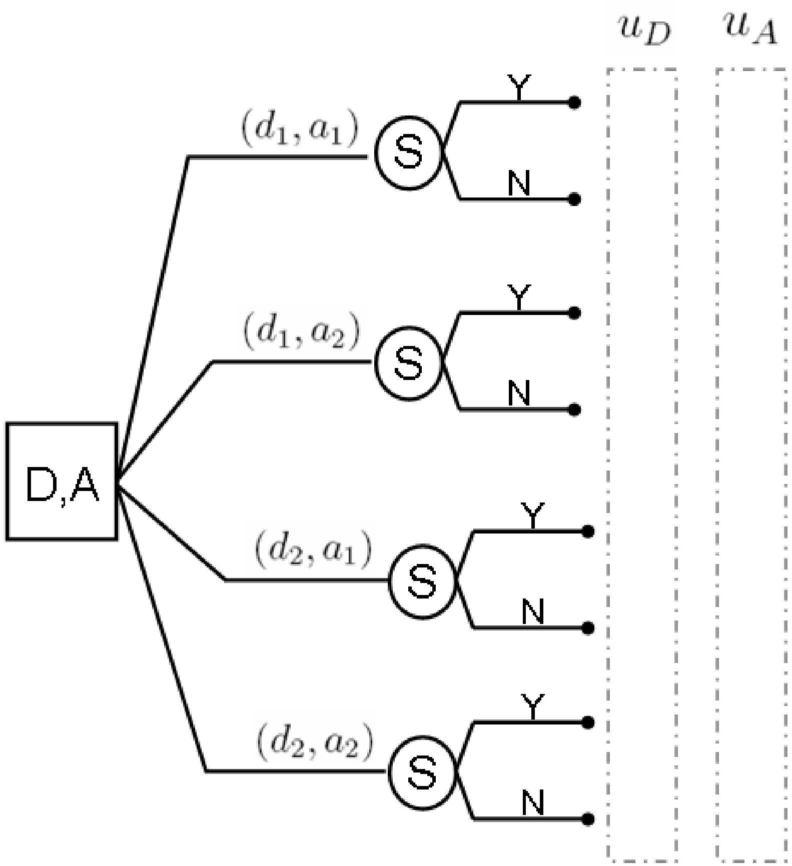
Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- **Simultaneous games**
- Auctions
- Security
- Intelligent interfaces
- Challenges

Simultaneous games

- Decisions are made without knowing each other's decisions





Game Theory Analysis

- Common knowledge
  - Each knows expected utility of every pair (d, a) for both of them
  - Nash equilibrium: (d*, a*) satisfying

$$\psi_D(d^*, a^*) \geq \psi_D(d, a^*) \ \ \forall\, d \in \mathcal{D}$$

$$\psi_A(d^*, a^*) \geq \psi_A(d^*, a) \ \ \forall\, a \in \mathcal{A}$$

- When some information is not common knowledge
  - Private information
    - Type of Defender and Attacker

$$\tau_D \in T_D \longrightarrow u_D(d, s, \tau_D) \quad p_D(S \mid d, a, \tau_D)$$

$$\tau_A \in T_A \longrightarrow u_A(d, s, \tau_D) \quad p_A(S \mid d, a, \tau_D)$$

  - Common prior over private information $\pi(\tau_D, \tau_A)$
  - Model the game as one of incomplete information

Bayes Nash Equilibrium

– Strategy functions

- Defender $d : \tau_D \to d(\tau_D) \in \mathcal{D}$
- Attacker $a : \tau_A \to a(\tau_A) \in \mathcal{A}$

– Expected utility of (d,a)

- for Defender, given her type $\psi_D(d(\tau_D), a, \tau_D) =$

$$= \int \left[\sum_{s \in S} u_D(d(\tau_D), s, \tau_D)\, p_D(S = s \mid d(\tau_D), a(\tau_A), \tau_D) \right] \pi(\tau_A \mid \tau_D)\, \mathrm{d}\tau_A$$

- Similarly for Attacker, given his type $\psi_A(d, a(\tau_A), \tau_A)$

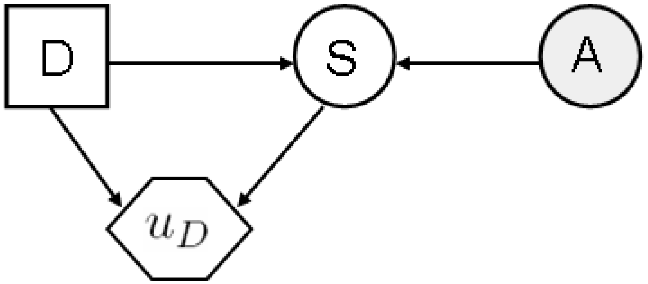
– Bayes-Nash Equlibrium (d*, a*) satisfying

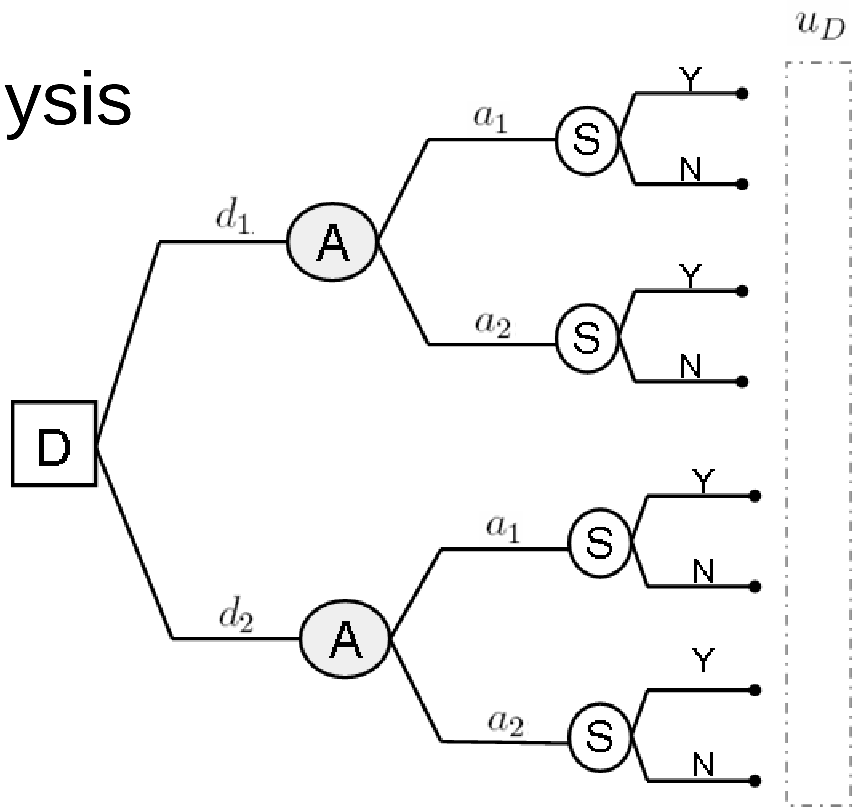
$$\psi_D(d^*(\tau_D), a^*, \tau_D) \geq \psi_D(d(\tau_D), a^*, \tau_D) \quad \forall\, d : \tau_D \to d(\tau_D)$$

$$\psi_A(d^*, a^*(\tau_A), \tau_A) \geq \psi_A(d^*, a(\tau_A), \tau_A) \quad \forall\, a : \tau_A \to a(\tau_A)$$

Supporting the Defender

- Defender's decision analysis



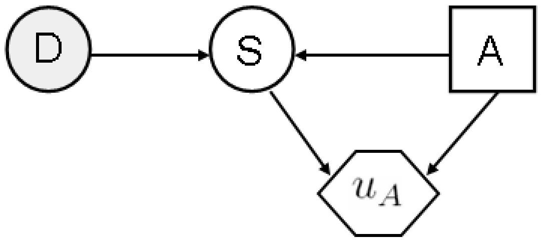


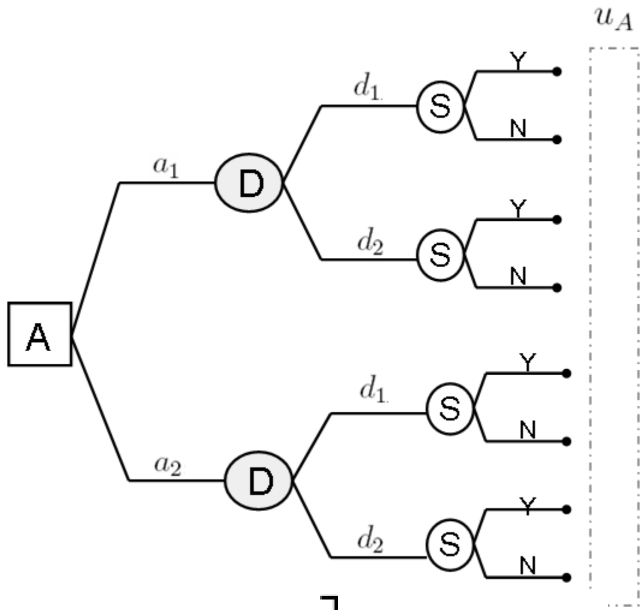
$$d^* = \operatorname{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} u_D(d,s)\, p_D(S = s \mid d, a) \right] \pi_D(A = a)$$

How to elicit it ??

Assessing $\pi_D(A = a)$

- Attacker's decision analysis as seen by the Defender





$$a^* = \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} u_A(a, s)\ p_A(S = s \mid d, a) \right] \pi_A(D = d)$$

$$(u_A, p_A, \pi_A) \sim (U_A, P_A, \Pi_A)$$

$$A \mid D \sim \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} U_A(a, s)\ P_A(S = s \mid d, a) \right] \Pi_A(D = d)$$

The assessment problem

- To predict Attacker's decision
  The Defender needs to solve Attacker's decision problem
  She needs to assess (u_A, p_A, π_A)

- Her beliefs about (u_A, p_A, π_A) are modeled through a probability distribution (U_A, P_A, Π_A)

- The assessment of $\Pi_A(D = d)$ requires deeper analysis
  - D's analysis of A's analysis of D's problem

- It leads to an infinite regress
  thinking-about-what-the-other-is-thinking-about…

Hierarchy of nested models

Repeat

Find $\Pi_{D^{i-1}}(A^i)$ by solving

$$A^i \mid D^i \quad \sim \quad \text{argmax}_{a \in \mathcal{A}} \sum_{d \in \mathcal{D}} \left[\sum_{s \in \{0,1\}} U_A^i(a,s)\ P_A^i(S = s \mid d, a) \right] \Pi_{A^i}(D^i = d)$$

$$\text{where } (U_A^i, P_A^i) \sim F^i$$

Find $\Pi_{A^i}(D^i)$ by solving

$$D^i \mid A^{i+1} \quad \sim \quad \text{argmax}_{d \in \mathcal{D}} \sum_{a \in \mathcal{A}} \left[\sum_{s \in \{0,1\}} U_D^i(d,s)\ P_D^i(S = s \mid d, a) \right] \Pi_{D^i}(A^{i+1} = a)$$

$$\text{where } (U_D^i, P_D^i) \sim G^i$$

$i = i + 1$

Stop when the Defender has no more information about utilities and probabilities at some level of the recursive analysis. K-level thinking

Opponent modeling

- Non strategic
- Nasheq
- Level-k
- Mirroreq
- Prospectmax

- Reconcile them through a mixture

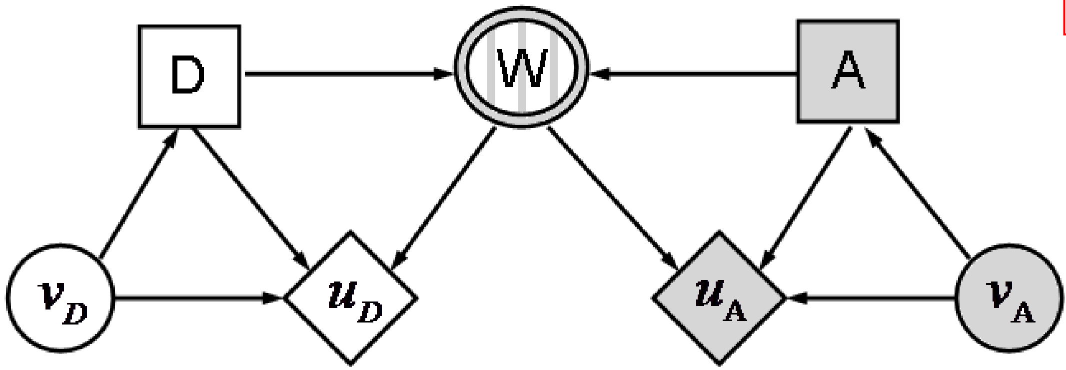
Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- **Auctions**
- Security
- Intelligent interfaces
- Challenges

Bidding in a two-person sealed-bid Auction

- Two sealed bids, the highest one wins
  - Simultaneous decision problem
- The standard Game Theory Analysis
  - D knows v_D but A does not: $\mathrm{p_A}(v_D)$
  - A knows v_A but D does not: $\mathrm{p_D}(v_A)$
  - Common knowledge assumption →

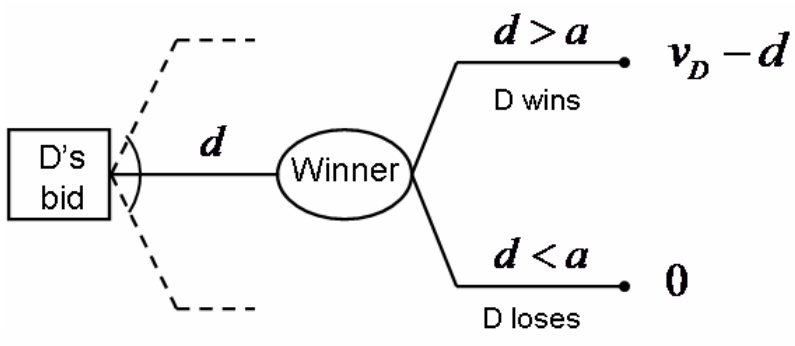
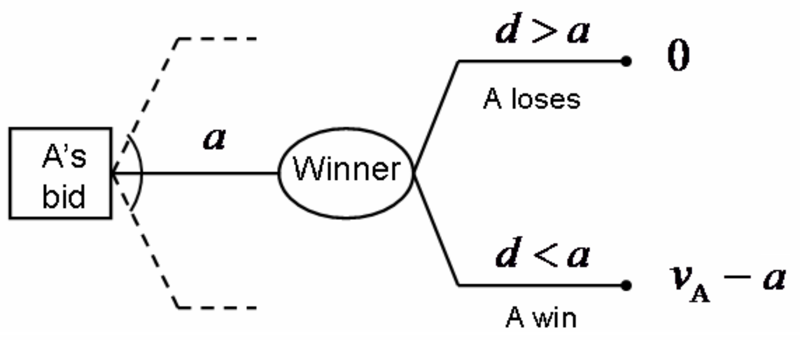
$$\mathrm{p_A}(v_D) = \mathrm{p}(v_D)$$
$$\mathrm{p_D}(v_A) = \mathrm{p}(v_A)$$



  - Bayesian Nash Eq. (Harsanyi)
  - Is it rational that players' beliefs about the opponent's object value will be disclosed??

Rothkopf (2007)

Supporting D

D's problem	D's analysis of A's problem
	
$\max_d u_D(v_D - d)\, \mathbb{P}_D(d > \underline{\underline{a}} \mid d)$	$\max_a \widehat{u}_A(\widehat{v}_A - a)\, \widehat{\mathbb{P}}_A(a > \underline{\underline{d}} \mid a)$
?? $(\widehat{u}_A, \widehat{v}_A, \widehat{\mathbb{P}}_A)$	$d \sim \widehat{\pi}_A$
	$\int_{-\infty}^{a} \widehat{\pi}_A(d)\, \mathrm{d}d$
	A's prob. of winning given his bid a

The assessment problem

- Assessment of $d \sim \widehat{\pi}_A$
- D's analysis of A's analysis of D's problem
  - It leads to a infinite analysis of previous analysis…
- Avoiding infinite regress
  - Available past statistical data (Capen et al, Keefer et al)
  - Expert knowledge
  - Non-informative distribution
  - Heuristic based elicitation (*)
- Heuristic elicitation $\widehat{\pi}_A(d)$
  - Identification of relevant variables in which A can base his assessment of D's bid $d \sim \widehat{\pi}_A$

Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- **Security**
- Intelligent interfaces
- Challenges

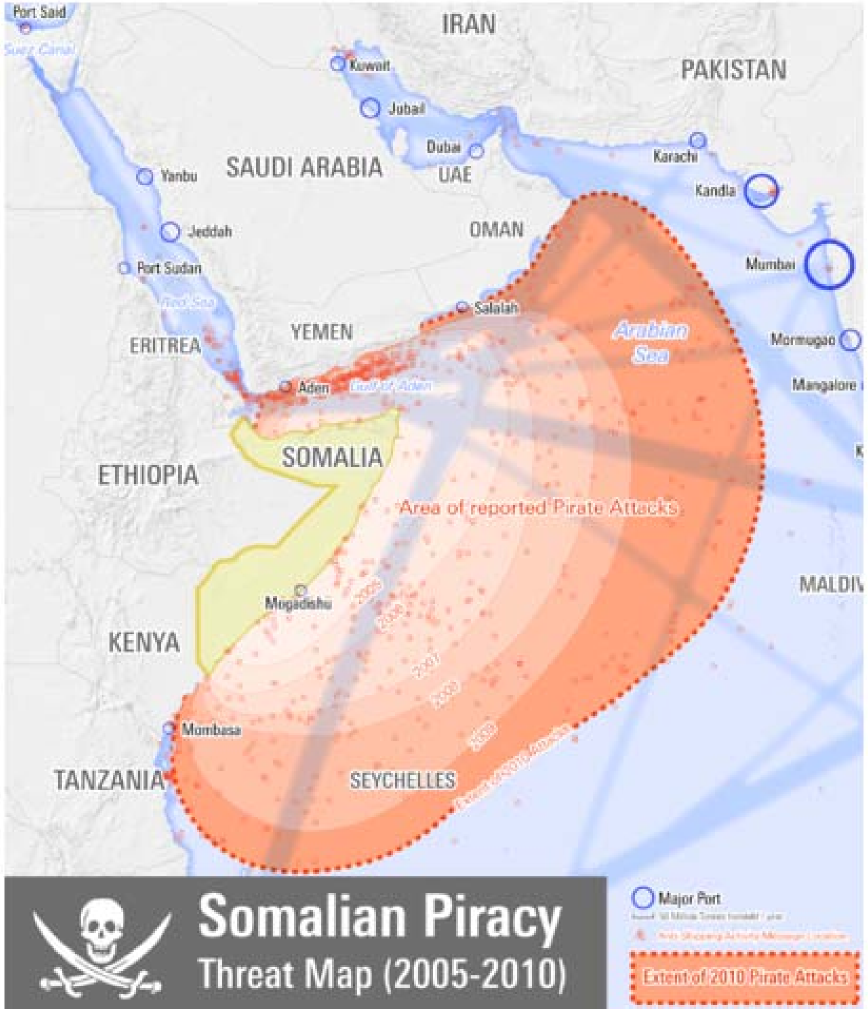
Security

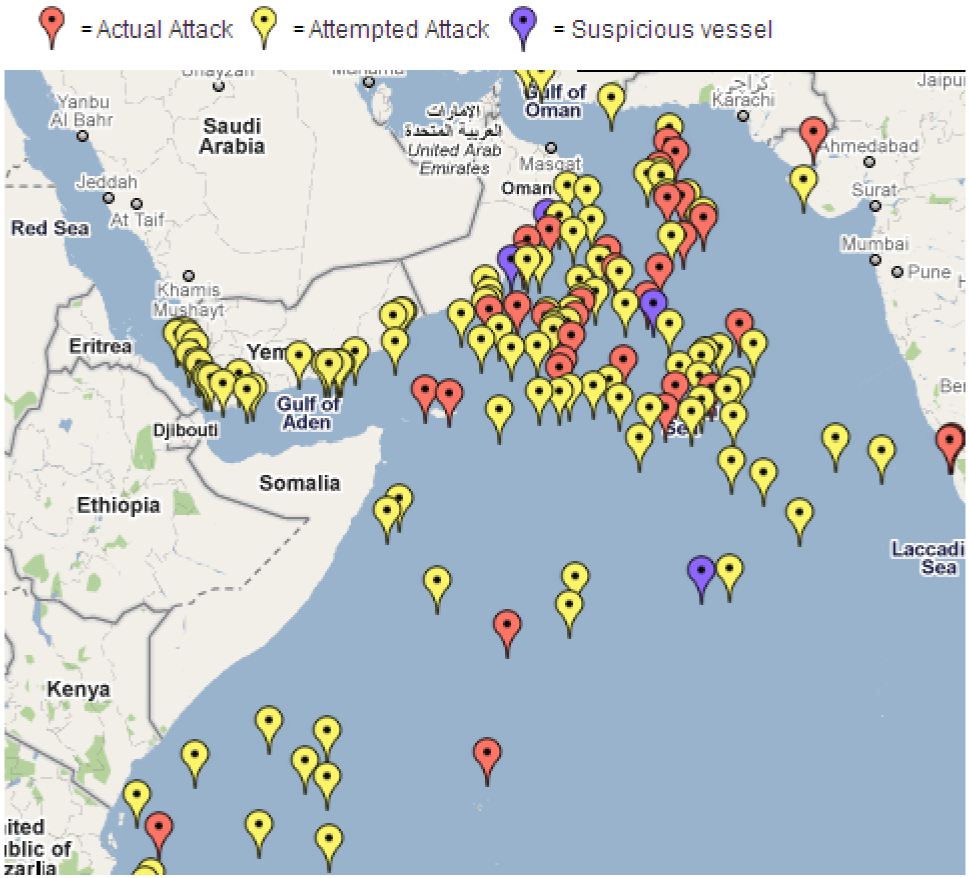
- One of ‘The World’s (23) Biggest Problems´ (Lomborg, 2008)
  - Arms proliferation
  - Conflicts
  - Corruption
  - Terrorism
  - Drugs
  - Money laundering

Security

- One of FP7 priorities. Horizon 2020
- SECONOMICS (2012-2015)
  - Anadulu Airport
  - Barcelona underground
  - National Grid, UK

Piracy in Somalia





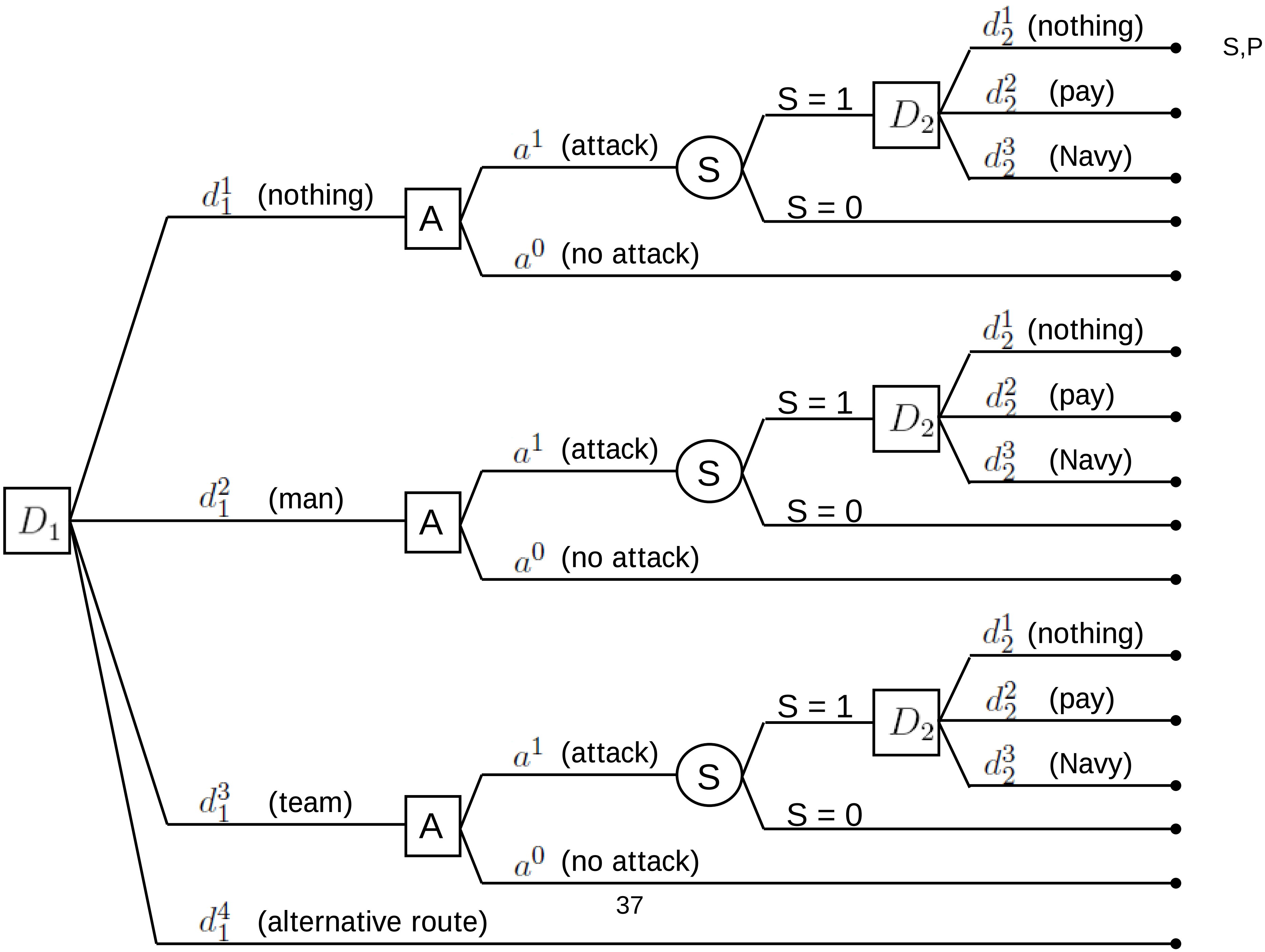
Piracy and armed robbery incidents reported to the IMB Piracy Reporting Centre 2011

The Defend–Attack–Defend model

- Two intelligent players
  - Defender and Attacker
- Sequential moves
  - First, Defender moves
  - Afterwards, Attacker knowing Defender’s move
  - Afterwards, Defender again responding to attack

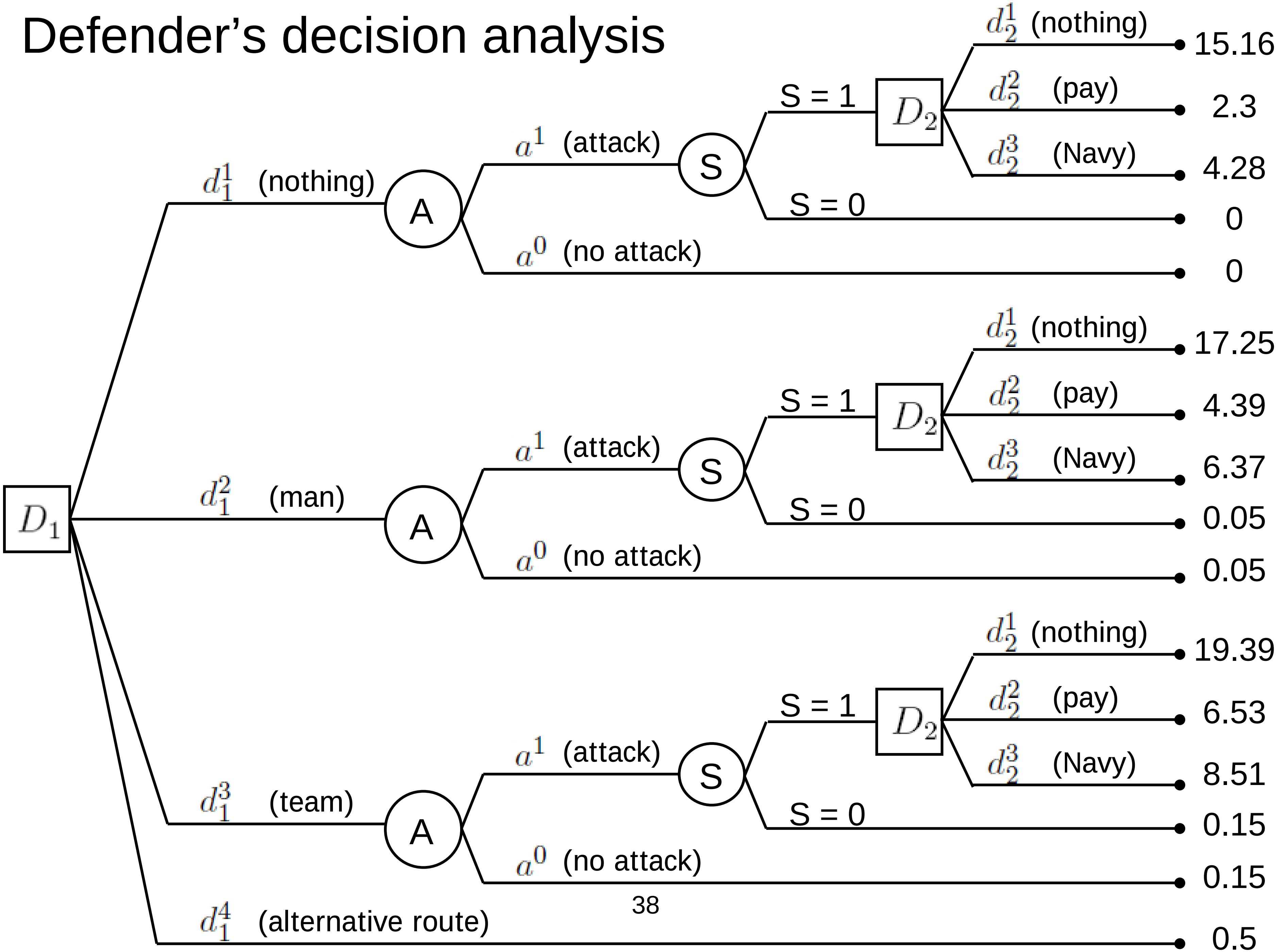
The Somali Pirates Case: Problem formulation

- Two players
  - Defender: Ship owner
  - Attacker: Pirates
- Defender first move
  - Do nothing
  - Private protection with an armed person
  - Private protection with a team of two armed persons
  - Go through the Cape of Good Hope avoiding the Somali coast
- Attacker's move
  - Attack or not to attack the Defender's ship
- Defender response to an eventual kidnapping
  - Do nothing
  - Pay the ransom
  - Ask the Navy for support to release the boat and crew



S,P
d_2^1 (nothing)
d_2^2 (pay)
d_2^3 (Navy)
S = 1
S = 0
D_2
a^1 (attack)
a^0 (no attack)
S
A
d_1^1 (nothing)
d_1^2 (man)
d_1^3 (team)
d_1^4 (alternative route)
D_1

Defender's decision analysis



ARA for Urban Security. Basics

- City divided into cells (i,j)
- Each cell has a value v_{ij}
- Actors
  1. Defender, D, Police. Aims at maintaining value
  2. Attacker, A, Mob. Aims at gaining value

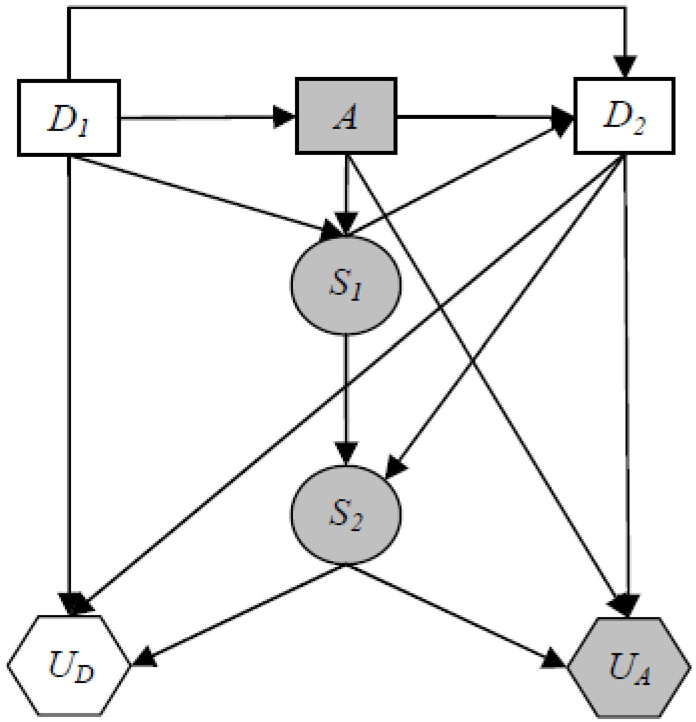
- D allocates resources to prevent $\sum_{ij} d_{ij}^1 \leq D_1$
- A performs attacks $\sum_{ij} a_{ij} \leq A$
- D allocates resources to recover $\sum_{ij} d_{ij}^2 \leq D_2$

  Plus other constraints

ARA for Urban Security. Basics

At each cell, a coupled influence diagram

Cell decision making coordinated by constraints on resources



ARA for Urban Security. Mob dynamics

Inicializar parámetros

Generar la estructura del ataque $\{d_1, a, s_1, d_2\}$ y $P_A^i(d_2 \mid d_1, a, s_1)$

1. Para el Atacante, desde i = 1,2,...,N repetir

   En el nodo S_2 y $\forall d_1, a, s_1, d_2$ factibles

   Generar $P_A^i(s_2 \mid s_1, d_2)$

   Obtener $\Psi_A^i(a, s_1, d_2, v) = \sum_{s_2} U_A^i(a, s_2, v) \prod_j P_A^i(s_j^2 \mid s_j^1, d_j^2)$

   En el nodo D_2 y $\forall d_1, a, s_1$ factibles

   Obtener $\Psi_A^i(d_1, a, s_1, v) = \sum_{d_2} \Psi_A^i(a, s_1, d_2, v) P_A^i(d_2 \mid d_1, a, s_1)$

   En el nodo S_1 y $\forall d_1, a$ factibles

   Generar $P_A^i(s_1 \mid d_1, a)$

   Obtener $\Psi_A^i(d_1, a, v) = \sum_{s_1} \Psi_A^i(d_1, a, s_1, v) \prod_j P_A^i(s_j^1 \mid d_j^1, a_j)$

   En el nodo A y $\forall d_1$ factible

   Obtener $(d_1, v) \rightarrow A_i^*(d_1, v) = \arg\max_{a \in A} \Psi_A^i(d_1, a, v)$

2. Aproximar $P_D(a \mid d_1)$ mediante

$$\hat{P}_D(a \mid d_1) = \frac{\#\{A_i^*(d_1, v) = a\}}{n}$$

3. Para el Defensor, hacer

   En el nodo S_2, $\forall d_1, a, s_1, d_2$

   Obtener $\psi_D(d_1, s_1, d_2, v) = \sum_{s_2} u_D(s_2, v) \prod_j p_D(s_j^2 \mid s_j^1, d_j^2)$

   En el nodo D_2, $\forall d_1, a, s_1$

   Obtener $\psi_D(d_1, s_1, v) = \arg\max_{d_2} \psi_D(d_1, s_1, d_2, v)$ y guardar $d_2^*(d_1, a, s_1)$

   En el nodo S_1, $\forall d_1, a$

   Obtener $\psi_D(d_1, a, v) = \sum_{s_1} \psi_D(s_1, d_2, v) \prod_j p_D(s_j^1 \mid d_j^1, a_j)$

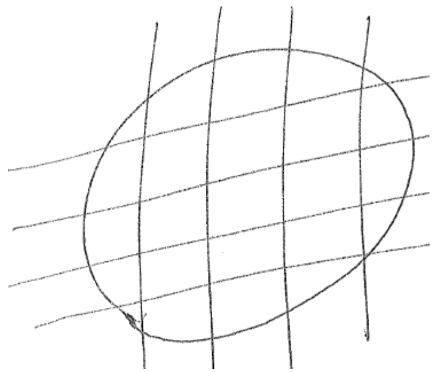
   En el nodo A, obtener $\forall d_1$

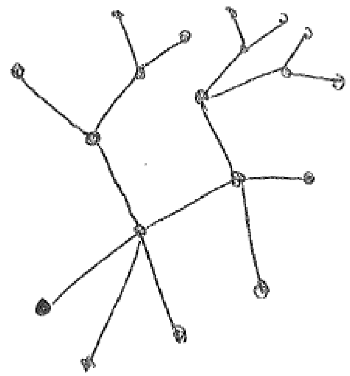
   Obtener $\psi_A(d_1, v) = \sum_a \psi_D(d_1, a, v) p_D(a \mid d_1)$

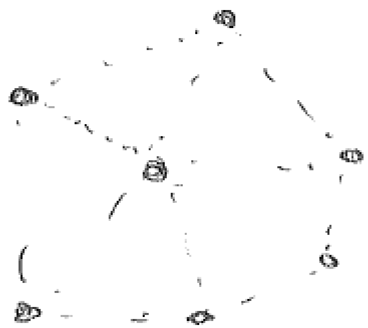
   En el nodo D_1

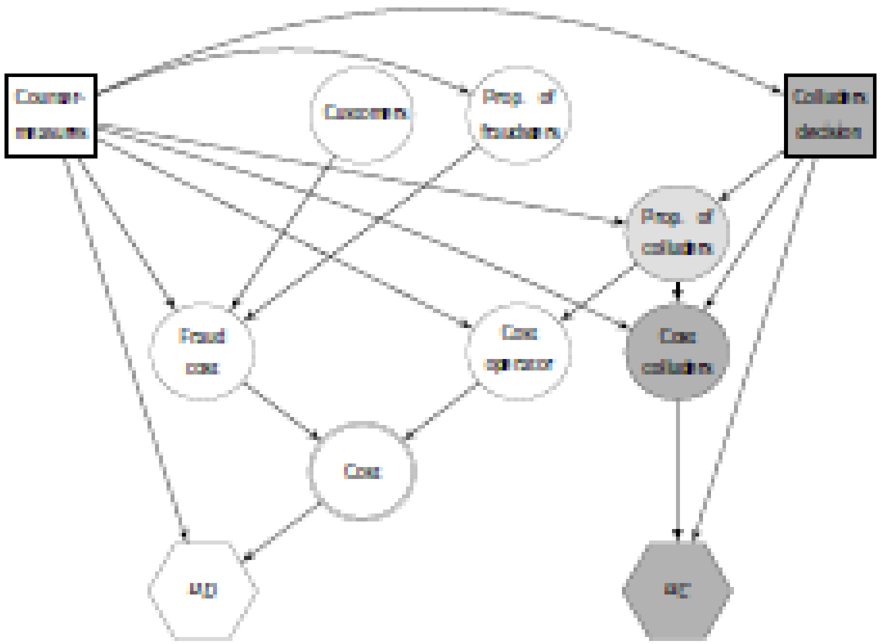
   Obtener $\psi_D(v) = \arg\max_d \psi_D(d_1, v)$ y guardar d_1^*

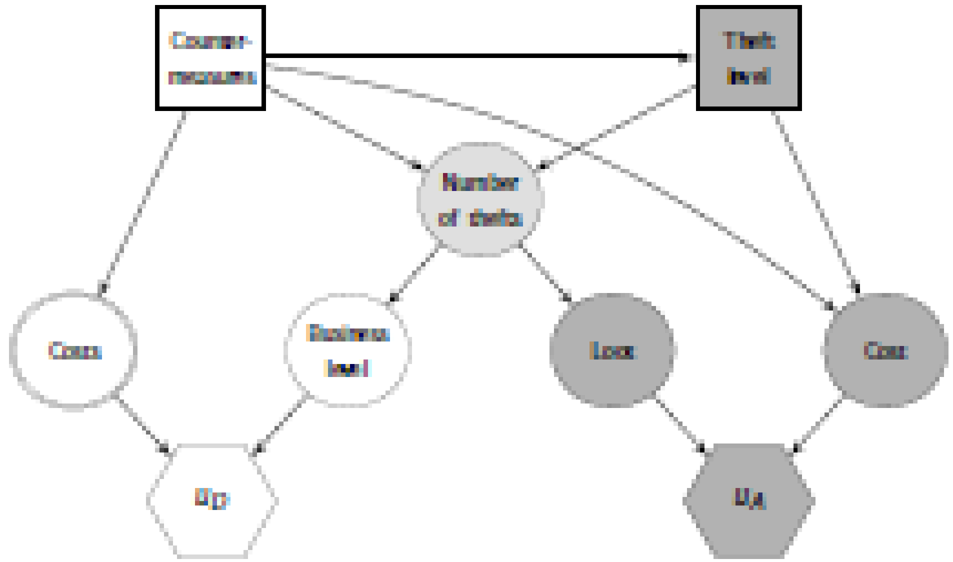
Security








Counter-measures
Customers
Prop. of fraudsters
Colluders decision
Prop. of colluders
Fraud cost
Cost operator
Cost colluders
Cost
u_D
u_C

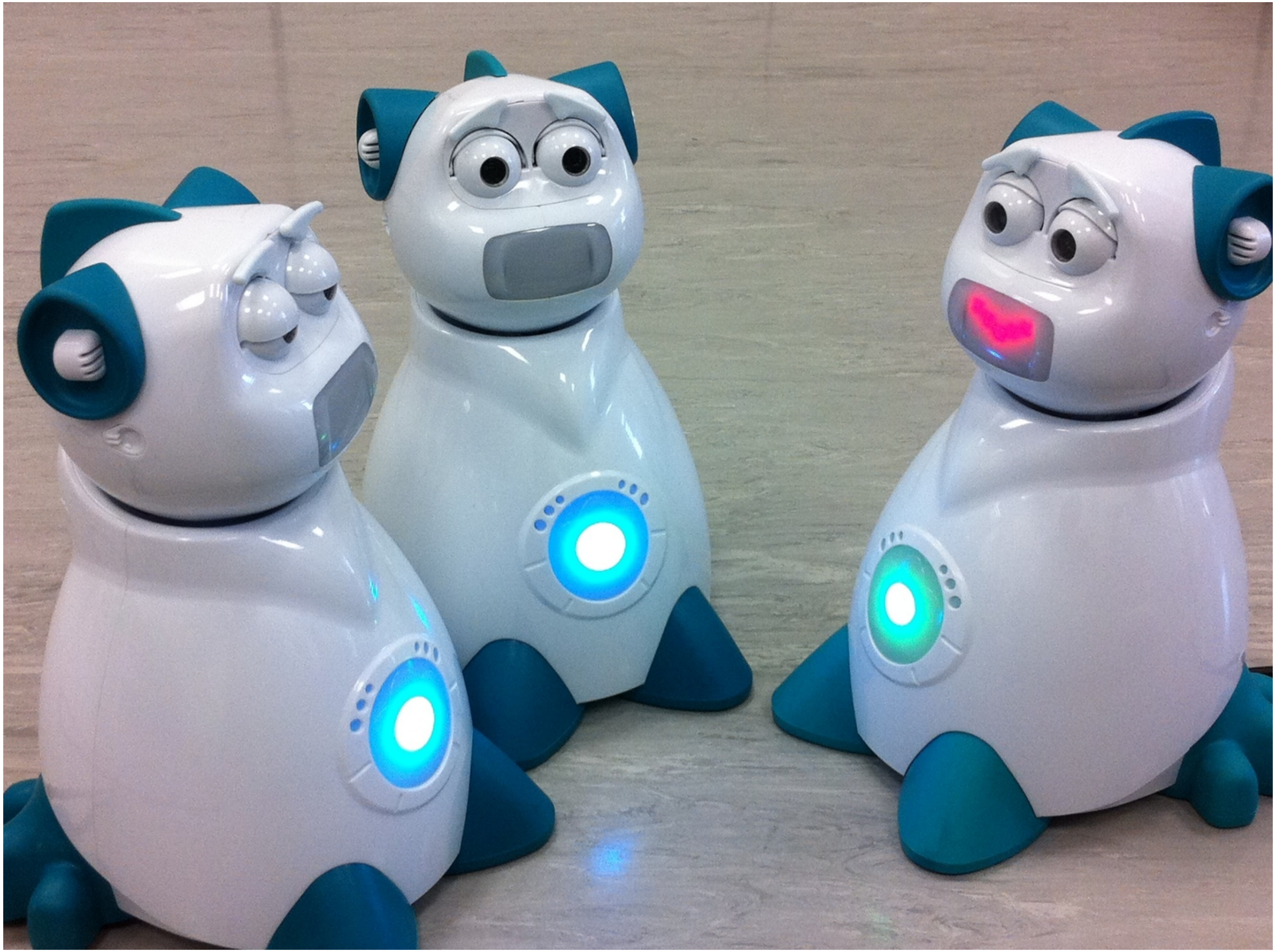

Counter-measures
Theft level
Number of thefts
Costs
Business level
Loss
Cost
u_D
u_A

Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- Security
- **Intelligent interfaces**
- Challenges

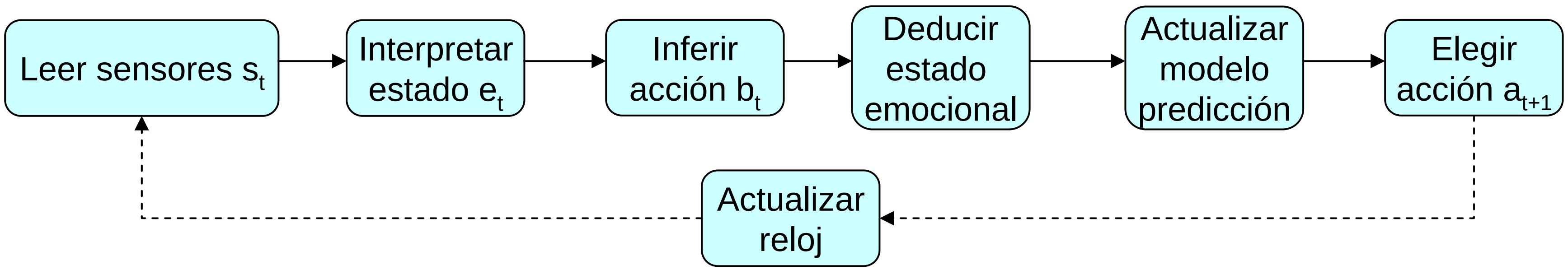
Problem

- An agent makes decisions in a finite set
- Has sensors providing information around it
- It relates with a user which makes decisions
- They're both within an environment which evolves (under the control of the user)

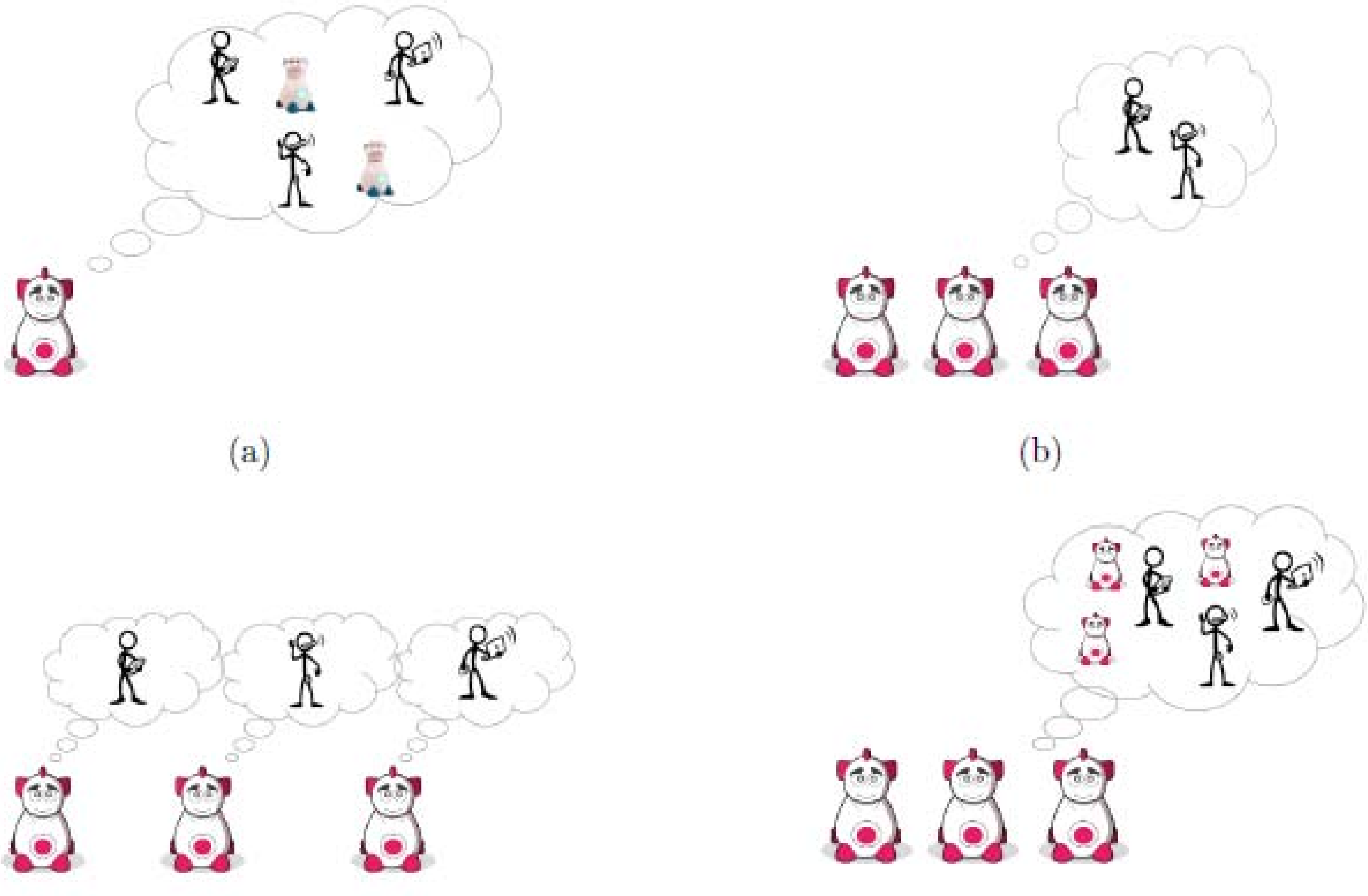


Basic framework

$$\max_{a_t \in \mathcal{A}} \psi(a_t) = \sum_{b_t, e_t} u(a_t, b_t, e_t) \times p(b_t, e_t \mid a_t, (a_{t-1}, b_{t-1}, e_{t-1}), (a_{t-2}, b_{t-2}, e_{t-2}))$$



Basic framework



Outline

- From risk analysis to adversarial risk analysis
- Motivation
- Sequential games
- Simultaneous games
- Auctions
- Security
- Intelligent interfaces
- **Challenges**

Challenges

- DA vs GT
  - A Bayesian prescriptive approach to support Defender against Attacker
  - Weaken common (prior) knowledge assumption
  - Analysis and assessment of Attacker' thinking to anticipate their actions assuming Attacker is a expected utility maximizer
  - Computation of her defense of maximum expected utility
  - What if the other not EU maximiser? Prospect theory, concept uncertainty

- Several simple but illustrative models
  - sequential D-A, simultaneous D-A, D-A-D, sequential DA with private information decision problems
  - What if
    - more complex dynamic interactions? (coupled IDs with shared nodes9
    - against more than one Attacker?
    - an uncertain number of Attackers?
    - several defenders? (risk sharing negotiations)

Challenges

- Implementation issues
  - Elicitation of a valuable judgmental input from Defender
  - Computational issues (optimization + simulation)
  - Augmented simulation
  - Parallel
  - Portfolio theory
  - Templates
  - K.level. The value of information
  - Computational environment

- Other applications
  - Revisiting Auctions
  - Revisiting Games
  - Cybersecurity
  - Adversarial signal processing
  - Network security

Discussion

- Multiple Defenders to be coordinated (risk sharing).
- Private security
- Multiple Attackers, possibly coordinated
- Various types of resources
- Various types of delinquency
- Multivalued cells. The perception of security (concern analysis)
- Multiperiod planning
- Time and space effects (Displacement of delicts)
- Insurance

- Networks with value only at nodes
- Networks with value at nodes and arcs

Discussion

- Educational environments
- Emotions and cooperativeness
- Multiperiod planning
- Mobility

Thanks!!!

david.rios@urjc.es

www.analisisderiesgos.com

www.aisoy.com