Expert judgement and adversarial problems

David Rios Insua
AXA-ICMAT Chair and Royal Academy

Delft, July 2017

Agenda

Adversarial problems

Adversarial Risk Analysis

ARA as a SEJ technique

Some advances in ARA in relation with EJNET

Thank you(s)

Adversarial problems

- Terrorism
- Business decisions: Auctions, Competitive marketing,…
- Cybersecurity
- …

One or more adversaries making decisions increasing our threats and affecting our results

Need to forecast what others will make

Reliability Analysis

How long will a system last under certain operational conditions?

Based on data and prior info…

- Make inferences about parameters present in lifetime models
- Make forecasts about lifetimes

To make decisions about replacement, maintenance, performance, design, configuration, …

Sometimes, several agents in scene: warranties, insurance, manufacturer(s)-consumer(s), regulator, security,…

Best HW/SW maintenance policy for a company ERP?

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

Forecast block reliabilities (and correlations)

Forecast system reliability

Design maintenance policies

Forecast their impact on reliability (performance, costs,…)

Optimal maintenance policy

Best HW/SW maintenance policy for a company ERP?

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

Forecast block reliabilities (and correlations)

Forecast system reliability

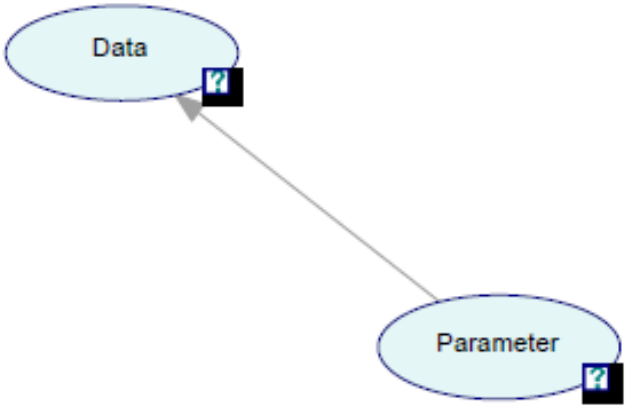
Design maintenance policies

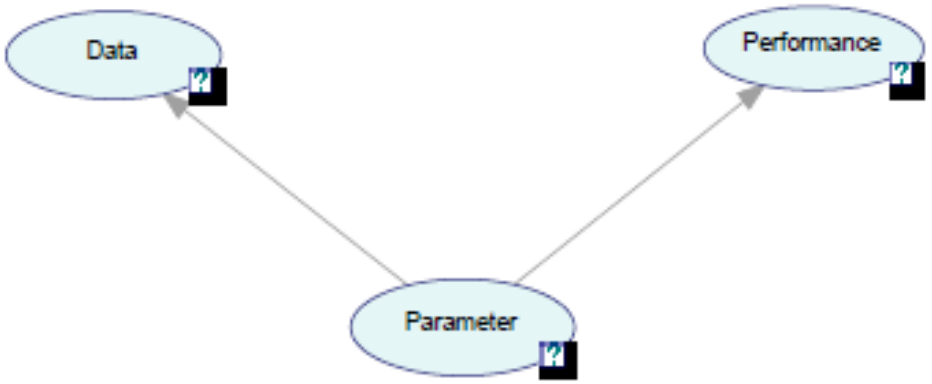
Forecast their impact on reliability (performance, costs,…)

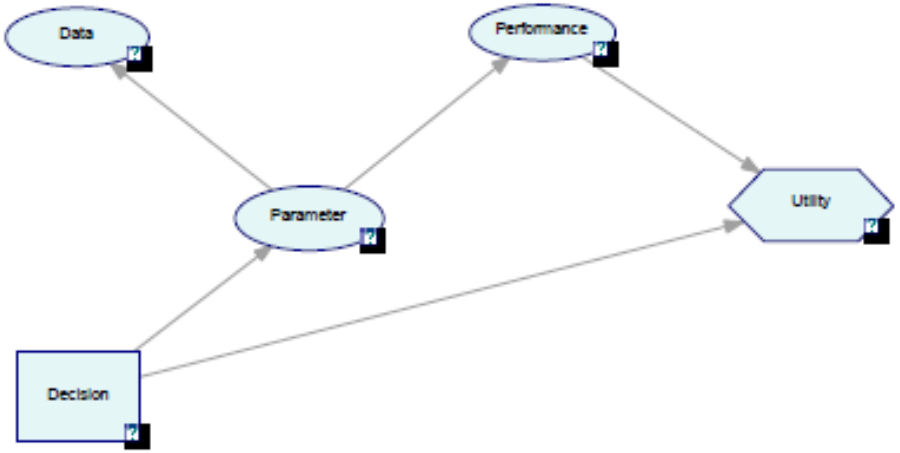
Optimal maintenance policy

NB: What happens with bad guys attacking our system?

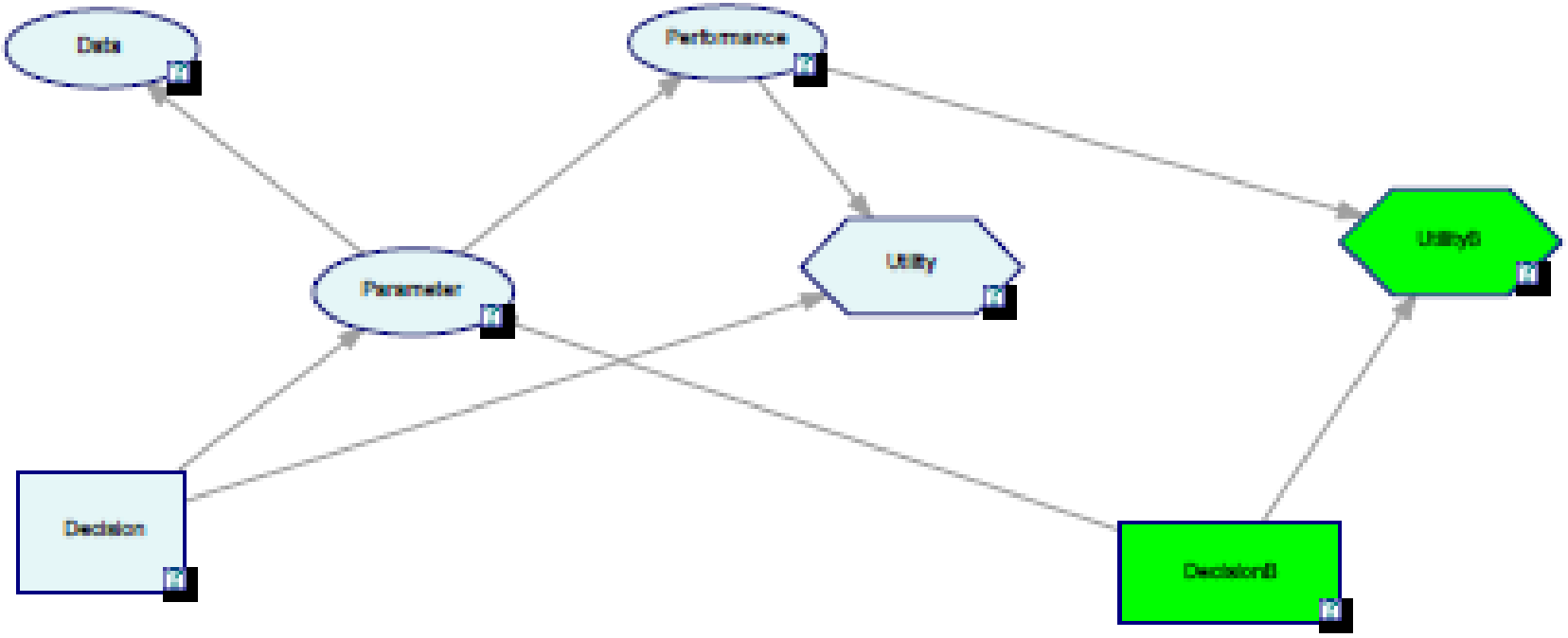
Reliability







Adversarial Reliability



Risk Analysis

What would be the impact over system performance of identified threats?

Based on data and prior info…

- Make forecasts of threat occurrence
- Make forecasts of threat impacts

To make risk management decisions

Sometimes, other agents in scene: security, cybersecurity, competitive marketing, social robotics, auctions,…

Best security resource allocation in a city?

City as a map with cells

Each cell has a value (multiattribute)

For each cell, a predictive model of delictive acts (COMPSTAT, PREDPOL,…)

Allocate security resources (given constraints)

For each cell predict impact of resource allocation

Optimal resource allocation

Best security resource allocation in a city?

City as a map with cells

Each cell has a value (multiattribute)

For each cell, a predictive model of delictive acts (COMPSTAT, PREDPOL,…)

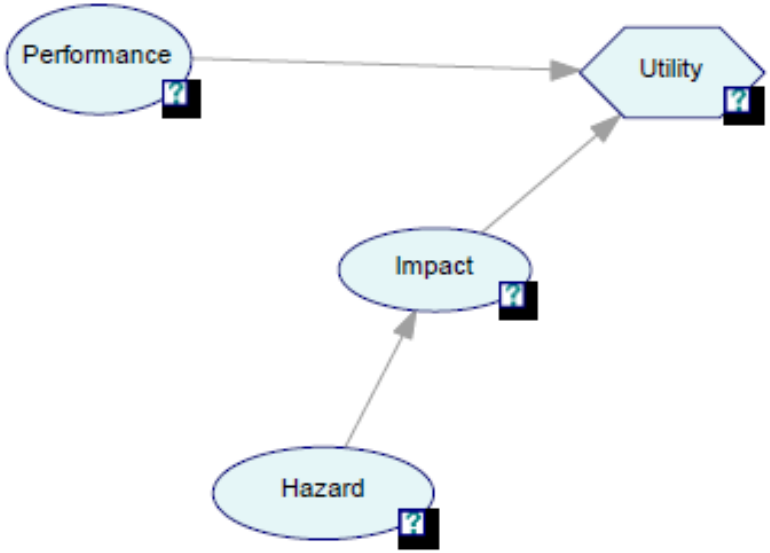
Allocate security resources (given constraints)

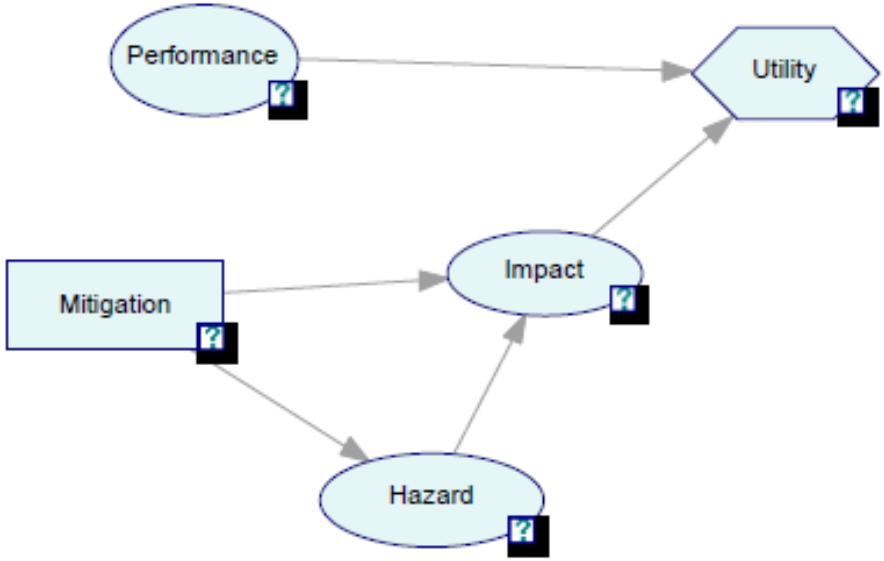
For each cell predict impact of resource allocation

Optimal resource allocation

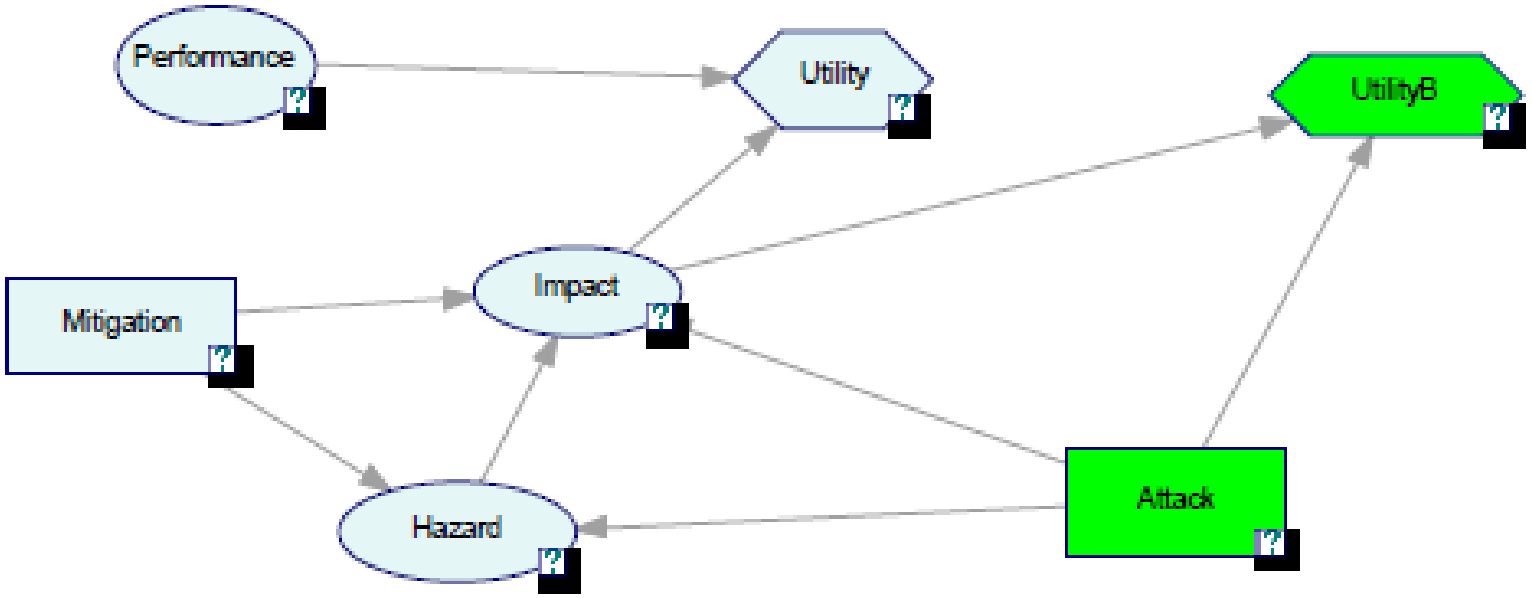
NB: The bad guys also operate intelligent and organisedly!!!

Risk Analysis





Risk Analysis



Agenda

Adversarial problems

Adversarial Risk Analysis

ARA as a SEJ technique

Some advances in ARA in relation with EJNET

Thank you(s)

Motivation

- 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 assumptions…
  - Decision analytic approaches. Forecasting the adversary action…
- Merrick, Parnell (2011) review approaches commenting favourably on ARA properly apportioning uncertainty

ARA

- 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

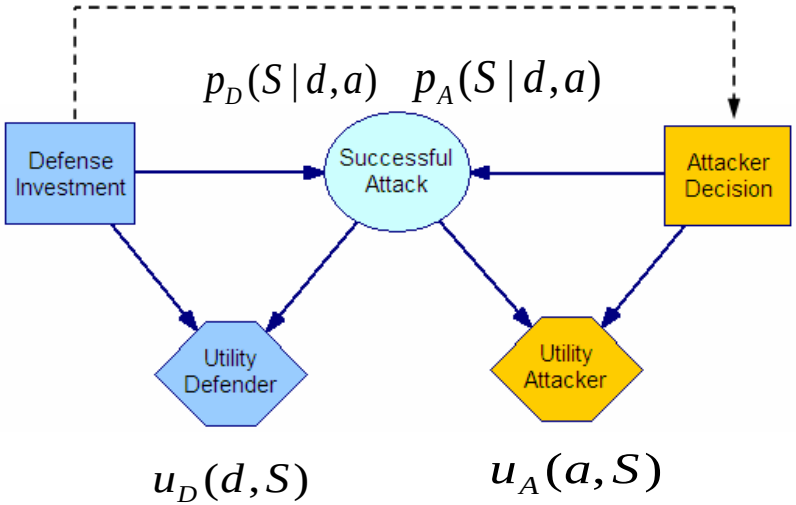
  (random, noninformative, level-k, heuristic, mirroring argument,…) vs (common knowledge)

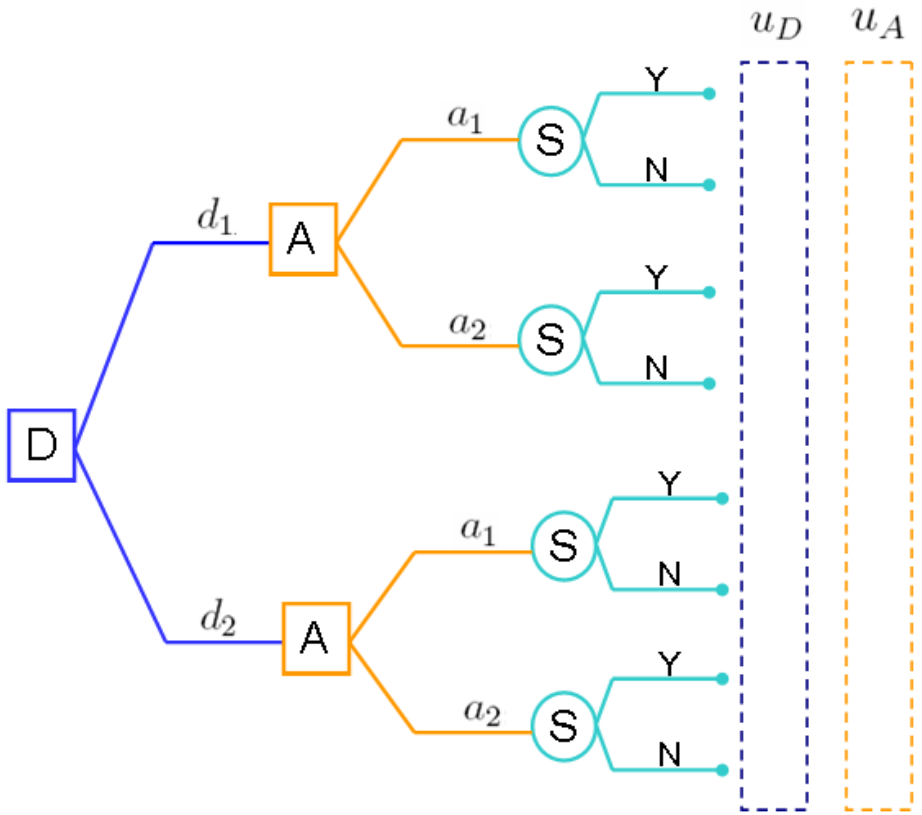
- Kadane, Larkey (1982), Raiffa (1982,2002)
- Lippman, McCardle (2012)
- Stahl and Wilson (1995) D. Wolpert (2012)
- Rothkopf (2007)
- MacLay, Rothschild, Guikema (2013,2014)

- Banks, Rios, DRI (2015)

Sequential DA game

- Two intelligent players
  - Defender and Attacker. D knows A's judgements
- Sequential moves
  - Def, then Attacker





Standard GT Analysis

Expected utilities at node S

$$\psi_D(d,a) = p_D(S=0|d,a)\ u_D(d,S=0)\ +\ p_D(S=1|d,a)\ u_D(d,S=1)$$

$$\psi_A(d,a) = p_A(S=0\mid d,a)\ u_A(a,S=0)\ +\ p_A(S=1\mid d,a)\ u_A(a,S=1)$$

Best Attacker's decision at node A

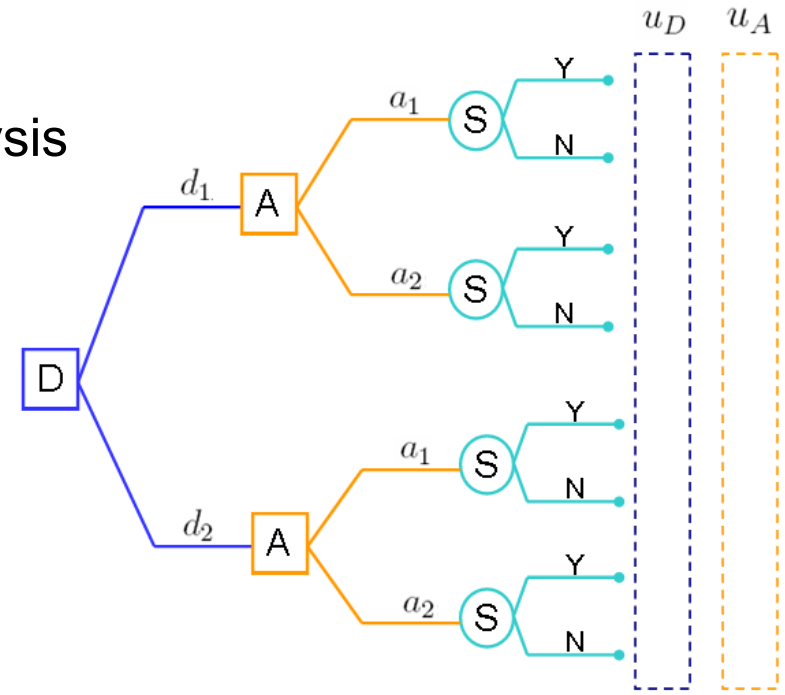
$$a^*(d) = \text{argmax}_{a\in\mathcal{A}}\psi_A(d,a)$$

Assuming Defender knows Attacker's analysis
Defender's best decision at node D

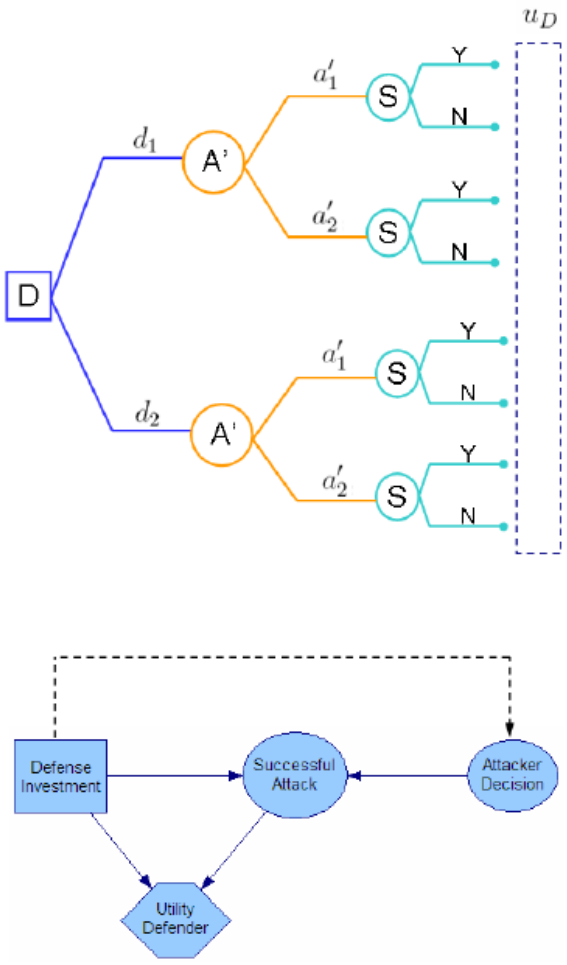
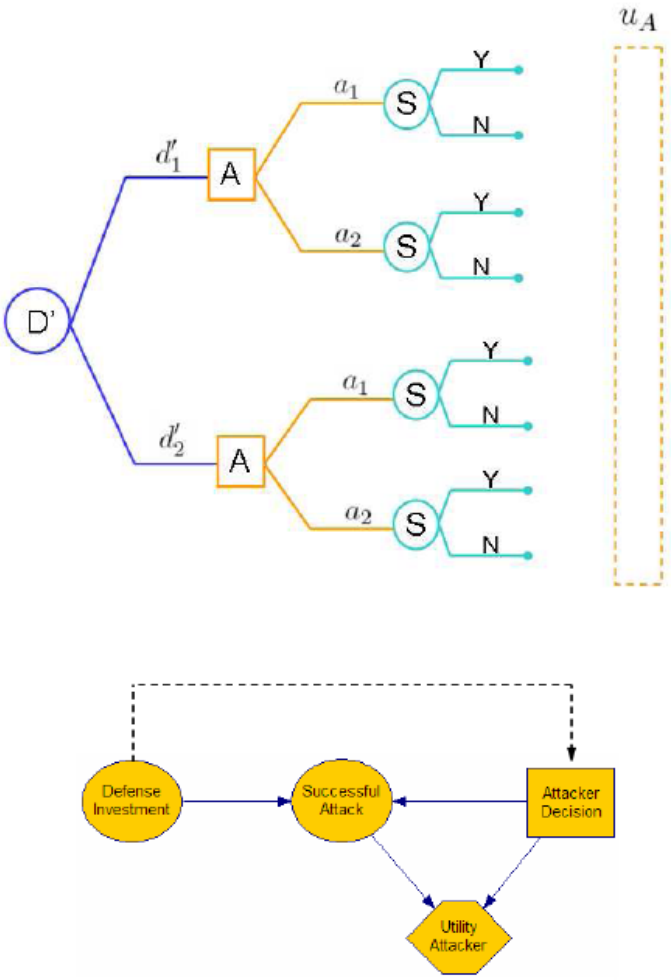
$$d^* = \text{argmax}_{d\in\mathcal{D}}\psi_D(d,a^*(d))$$

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

Nasheq. Subgame perfect equilibrium

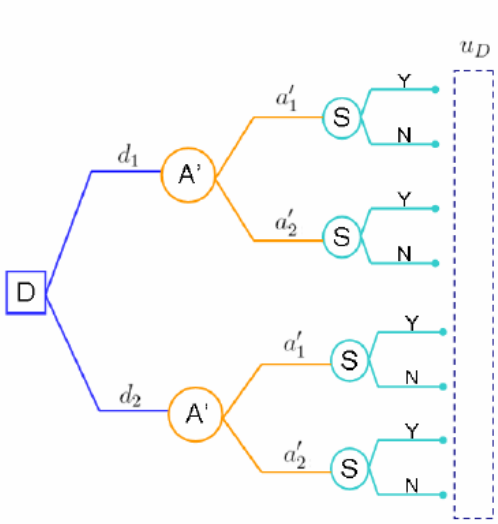


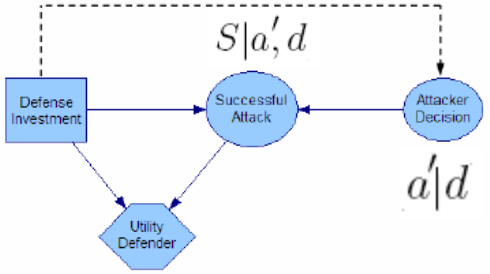
Supporting the Defender

Defender problem	Defender's view of Attacker problem
	

Supporting the Defender

Defender problem	Defender's solution





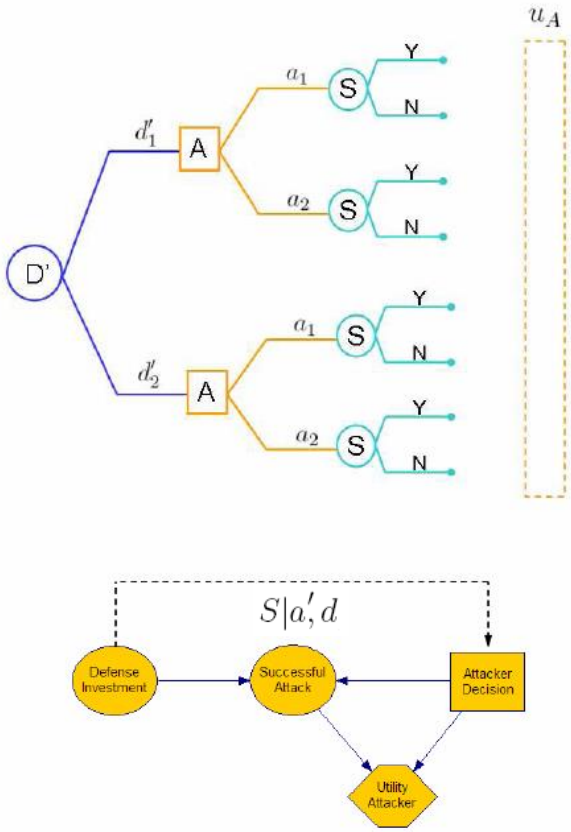
$$\psi_D(d, a') = u_D(d, S = Y)\, p_D(S = Y \mid X_D = d, X'_A = a') + u_D(d, S = N)\, p_D(S = N \mid X_D = d, X'_A = a')$$

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

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

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 $p_D(a'\|d)$
	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> $\hat{p}_D(a\|d) \approx n^{-1} \sum_i \#\{a = \text{argmax}_{x \in \mathcal{A}} \hat{\psi}_A^i(x, d)\}$ where $\hat{\psi}_A^i \sim \widehat{\Psi}_A,\ i = 1, \ldots, n$

Sequential D-A

1. Assess (p_D, u_D) from the Defender

2. Assess $F = (P_A, U_A)$, describing the Defender's uncertainty about (p_A, u_A)

3. For each d, simulate to assess $p_D(A|d)$ as follows:

   (a) Generate $(p_A^i, u_A^i) \sim F$, $i = 1, \ldots, n$

   Solve $a_i^*(d) = \text{argmax}_{a \in \mathcal{A}} \; \psi_A^i(d, a)$

   (b) Approximate $\hat{p}_D(A = a|d) = \#\{a = a_i^*(d)\}/n$

4. Solve the Defender's problem

$$d^* = \text{argmax}_{d \in \mathcal{D}} \; \psi_D(d, a_1)\, \hat{p}_D(A = a_1|d) \; + \; \psi_D(d, a_2)\, \hat{p}_D(A = a_2|d)$$

Simultaneous and beyond gets more complicated!!!

Agenda

Adversarial problems

Adversarial Risk Analysis

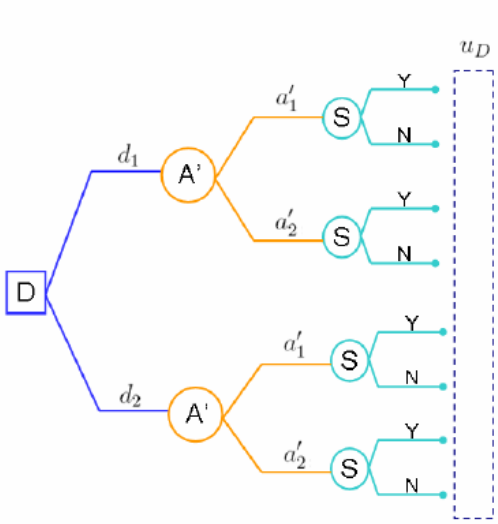
ARA as a SEJ technique

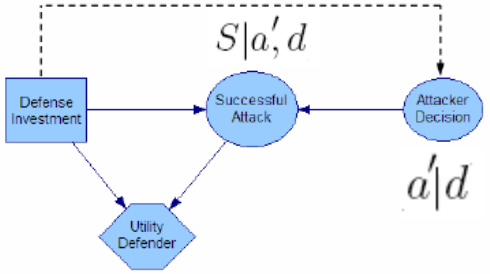
Some advances in ARA in relation with EJNET

Thank you(s)

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

Fermitisation (Tetlock)

- Extension of the conversation

Decompose a complex probability into probabilities simpler to assess who are then combined by total probability formula

Fermitisation (Tetlock)

- Extension of the conversation

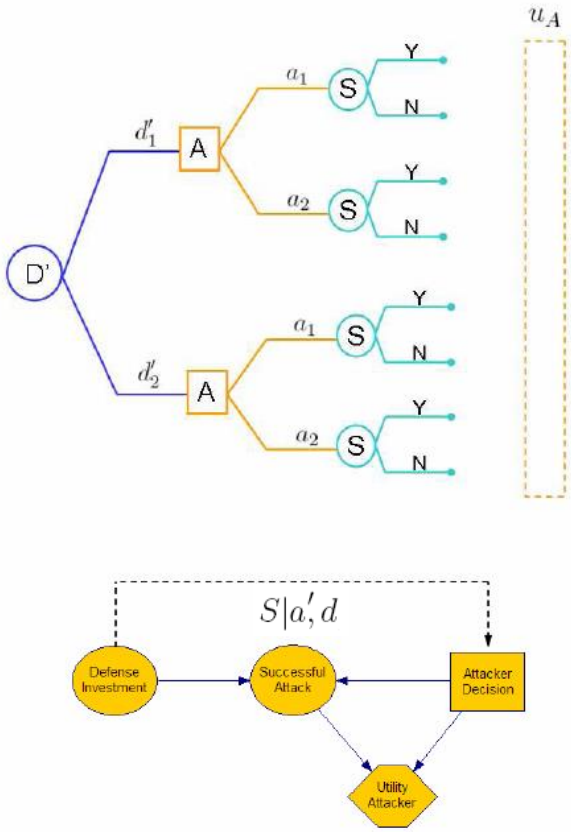
Decompose a complex probability into probabilities simpler to assess who are then combined by total probability formula

- ARA

Decompose a complex probability into probabilities simpler to assess who are then combined by ***maximising random expected utilities***

Decision Analysis!!!!

Supporting the Defender: The assessment problem

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> $\hat{p}_D(a\|d) \approx n^{-1} \sum_i \#\{a = \text{argmax}_{x \in \mathcal{A}} \hat{\psi}_A^i(x, d)\}$ where $\hat{\psi}_A^i \sim \widehat{\Psi}_A,\ i = 1, \ldots, n$

Agenda

Adversarial problems

Adversarial Risk Analysis

ARA as a SEJ technique

Some advances in ARA in relation with EJNET

Thank you(s)

ARA EJNET Advances

Conceptual

Methodological

Foundational

Computational

Applied

Conceptual. GT solutions not robust and SARA

- GT solutions robust. A Flat Maxima Principle
- GT solutions actually not robust!!!

If GT solution robust, STOP.

Else, ARA.

If ARA robust, STOP

Else, gamma-minimax et al

Conceptual. Opponent modeling

- Aleatory uncertainty. Risk Analysis
- Epistemic uncertainty. Model mixing

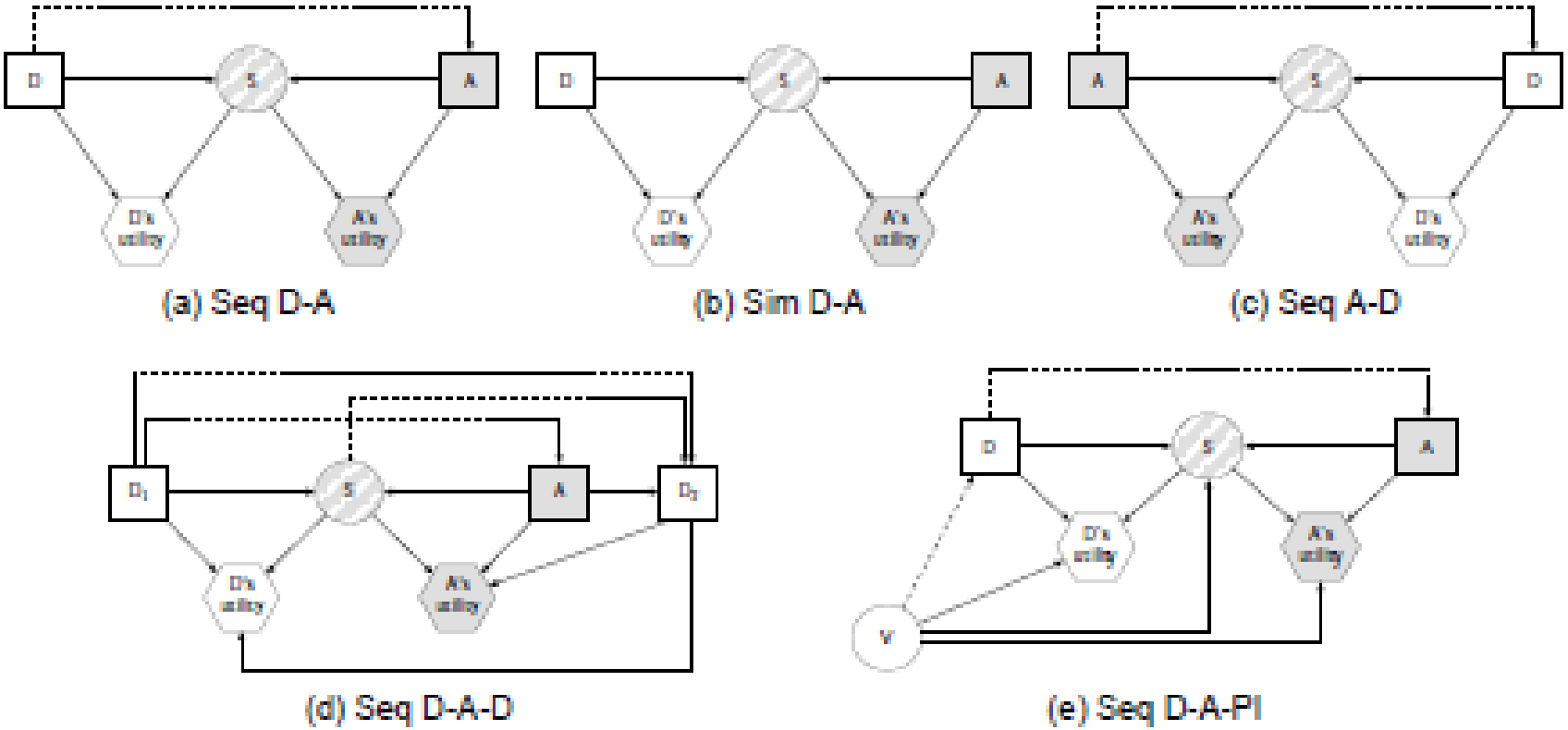
- Concept uncertainty

Reconcile various concepts through a mixture

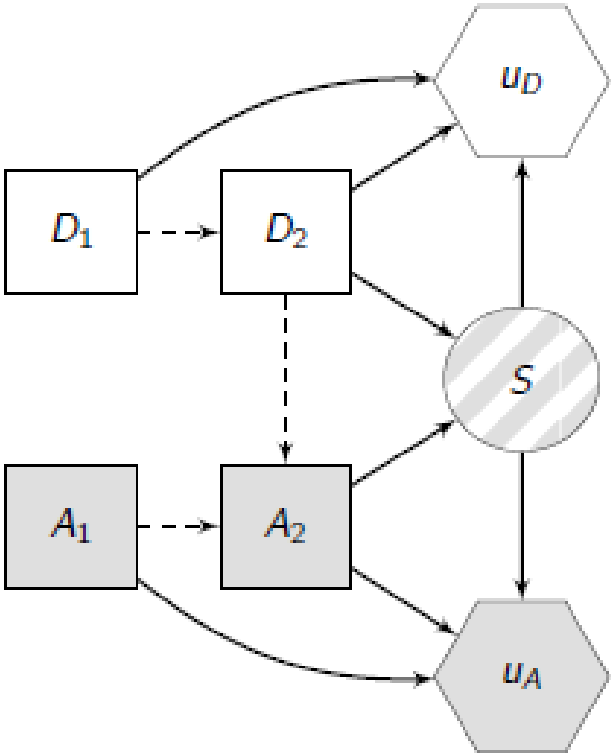
Opponent modeling

- Non strategic
  - ‘Against Nature’. Multi-Dir. Markov memory models. Fictitious play
- Nasheq
  - Opponent seeks a Nash eq.
- Level-k
  - Hierarchy. Stop when no more info available. Noninformative
- Mirroreq
  - Consistency condition for Defender beliefs.
- Prospectmax
  - Maximises a prospect theory functional
- ….

Computational. Beyond the templates

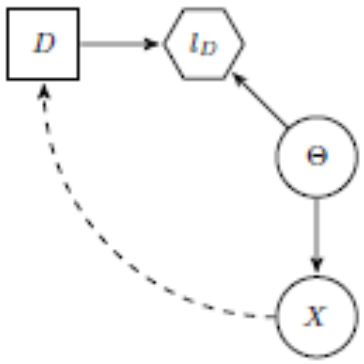


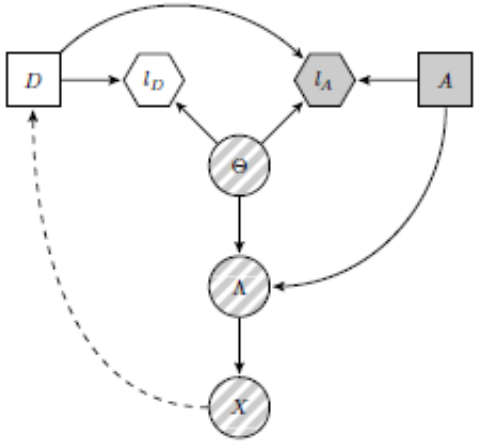
More general interactions



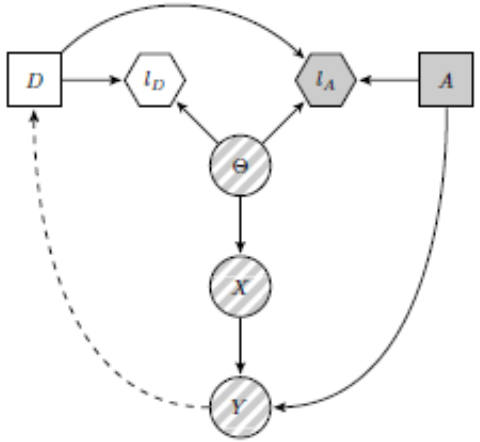
A method using the relevance graph

Foundational.
Adversarial Statistical Decision Theory



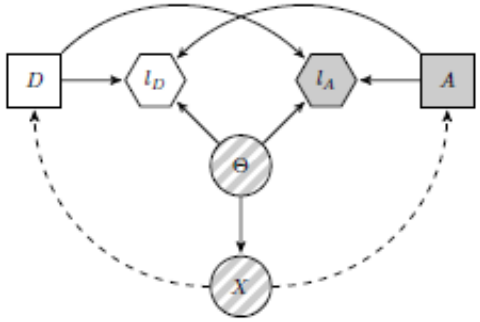


(a) Structural attacker



(b) Data-fiddler attacker

A Point Estimation
A Inter. Estimation
A Hypothesis Test.
A Prediction
A Classification
A Machine Learning
….
All things adversarial???



(c) Simultaneous ASDT problem

Applied: Case Studies and Applications

Problem	Defender	Attacker	Specificities	Template
ATC protection	Airport authority	Terrorist	Single site	D-> A
Piracy	Ship owner	Pirates	Single site	D- >A - > D
Metro	Operator	Pickpock Fare evasion	Multisite Multiattack, Cascade	D->A
Urban security	Police	Mob	Multisite spatial	D->A->D
Train	DoT, DoD	Terrorist	Multisite network	D->A->D
SME IS	Company	Competitor	Cyber, Integrated with RA. Cyberins	D->A
Oil rig cybercontrolled	Oil company	Sponsored hackers	Cyber, Multiattack	D->A->D
CI	Owner	Terrorist	Multistage	General
Social Robotics	Robot	User	Multistage, Emotions	D->A->D

Acceptance sampling, Spam detection, Fraud detection, Energy Risk, Defence vs UAV, Cybersec,…

Methodological. The ARA cycle

1. Structure problem
   - Underlying topological structure
     (single site, spacial, network, multiple site,…)
   - Determine Defenders and eventual coordination (single, multiple uncoordinated, multiple coordinated)
   - Determine Attackers, rationality style and eventual coordination
     (single, uncoordinated, cascade, coordinated,…)
   - Relevant template for each attacker and site
     (D A, D->A, D->A->D, BAID,…)
   - Expand templates for additional uncertainties
   - Determine resources and resource constraints

The ARA cycle

1. Structure problem
2. Assess problem
   - Determine Defender's own objectives, utilities, probabilities.
   - Determine Attacker's objectives, (random) utilities, (random) probabilities, as required
3. Solve problem
   - Simulate attacker problem to forecast actions
   - Optimise defender problem for optimal resource allocation
   - Sensitivity analysis
   - Communicate

Agenda

Adversarial problems

Adversarial Risk Analysis

ARA as a SEJ technique

Some advances in ARA in relation with EJNET

Thank you(s)

David Banks, Jesus Rios, Refik Soyer, Fabrizio Ruggeri, Jorge Ortega, Ahti Salo, Juho Roponen, Dani Rasines, Vesela Radovic, Cesar Alfaro, Javi Gomez, Aitor Couce, Siv Houmbd, Wolter Pieters, Roi Naveiro, Tinu Adebanji, Alberto Redondo,….

Bedankt EJNET!!!

Bedankt Roger!!!