An Adversarial Risk Analysis Approach to Fraud Detection

J. Cano[1] D. Ríos Insua[2]

[1]URJC

[2]ICMAT-CSIC, Spain

20th IFORS. Barcelona. July 15, 2014

Outline

SECONOMICS

A framework for risk analysis

A framework for adversarial risk analysis

A framework for risk analysis and adversarial risk analysis

Case study: fighting fare evasion

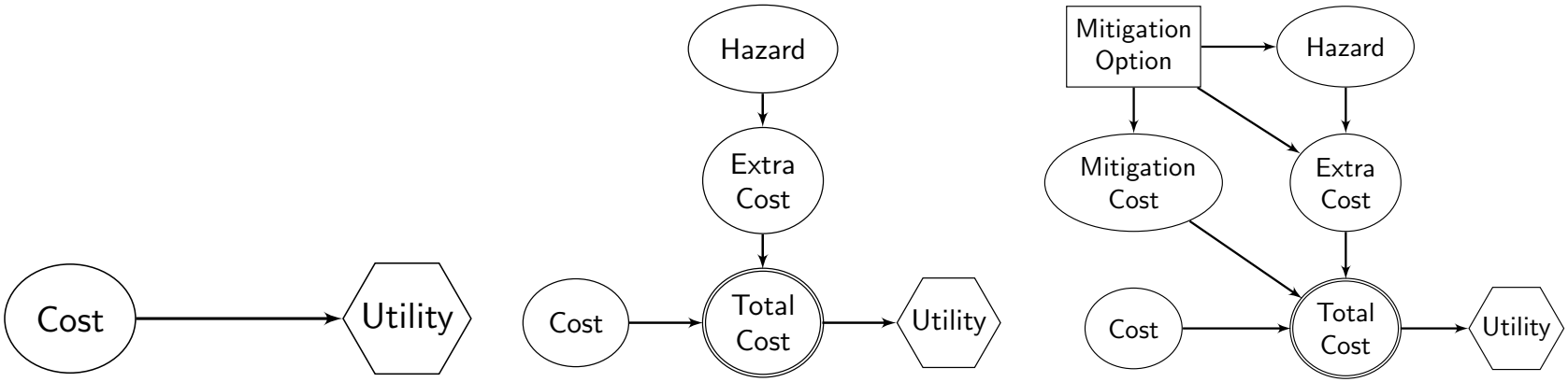
General overview

SECONOMICS

- Risk analysis $\rightarrow$ methodology to mitigate negative effects of threats that may harm system performance.
- Adversarial risk analysis $\rightarrow$ expands RA to deal with intelligent intentional adversaries.
- Application in fraud detection in relation with access to a paid facility.

1. A framework for risk analysis

Risk analysis influence diagrams and expected utility

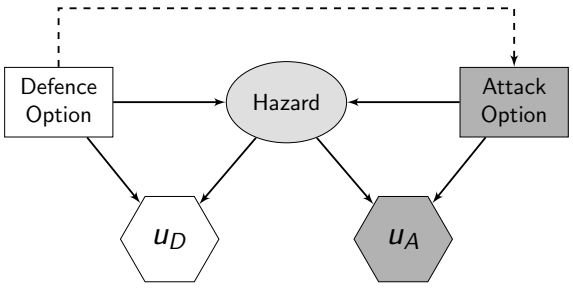


Basic With risk assessment With risk management

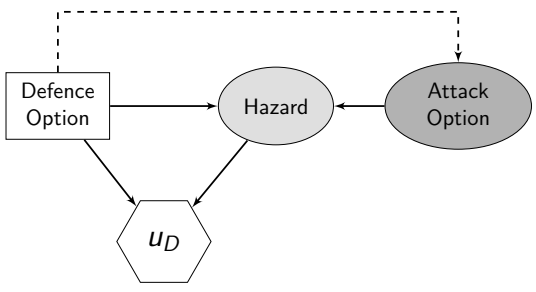
$$\psi = \int u(c)\pi(c)\mathrm{d}c, \quad \psi_r = \sum_{j=0}^{n} q_j \int u(c)\pi_j(c)\,\mathrm{d}c, \quad \psi_m = \max_{m\in\mathcal{M}} \sum_{j=0}^{n} q_j(m) \int u(c)\pi_j(c|m)\,\mathrm{d}c$$

2. A framework for adversarial risk analysis

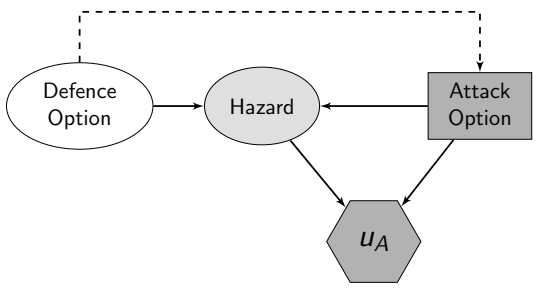
Sequential Defend-Attack model



Coupled



Defender's problem



Attacker's problem

Solving strategy

SECONOMICS

- Defender aims at finding optimal defense d^*.
  - Consequences evaluated through utility $u_D(d, s)$

$$\psi_D(d|a) = \int u_D(d, s)\, p_D(s|d, a)\, \mathrm{d}s.$$

- Suppose Defender able to assess $p_D(a|d)$.
- Then, she can compute

$$\psi_D(d) = \int \psi_D(d|a)\, p_D(a|d)\, \mathrm{d}s.$$

and solve

$$d^* \longleftarrow \max_{d \in \mathscr{D}} \quad \psi_D(d).$$

Assessment of Attacker's intentions

SECONOMICS

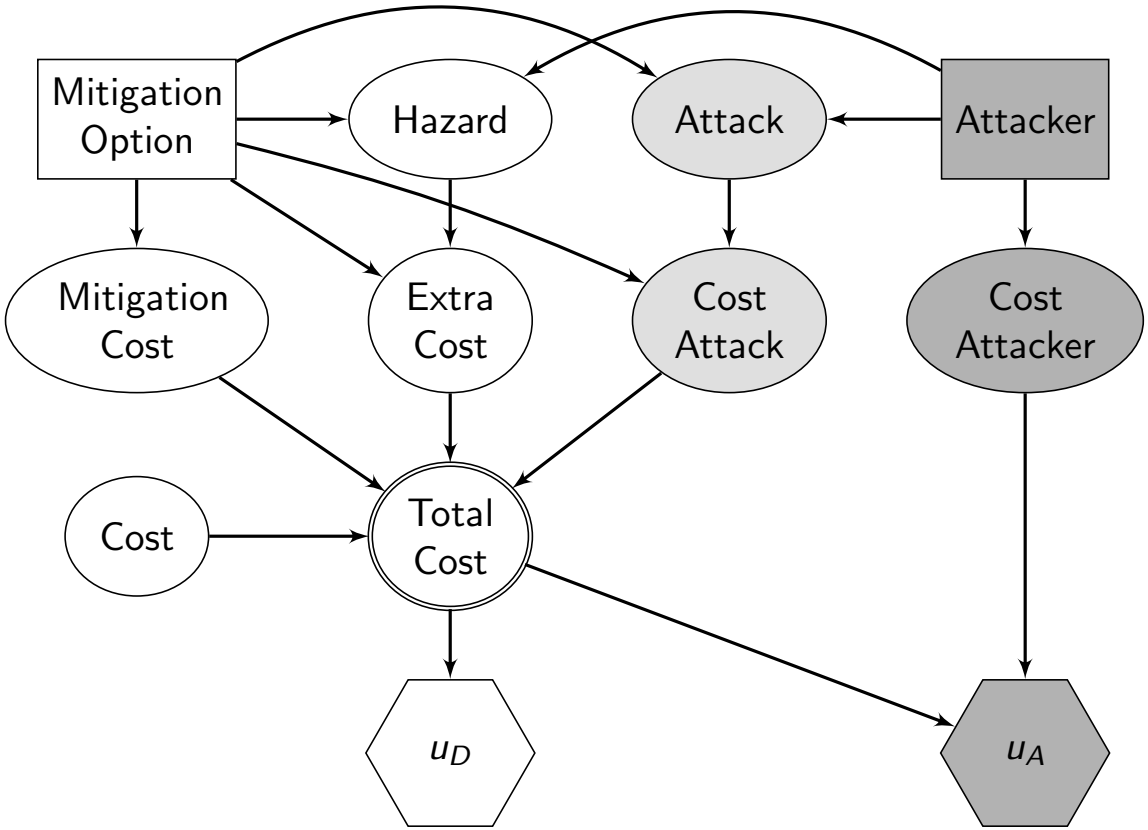
- To obtain $p_D(a|d)$, solve Attacker's problem (E.U. max.)

$$a^*(d) = \arg\max_{a \in \mathscr{A}} \int u_A(a,s)\, p_A(s|d,a)\, \mathrm{d}s.$$

- Defender lacks knowledge $\left(u_A(\cdot), p_A(s|\cdot)\right) \to \left(U_A, P_A\right)$.
- Approximate $\widehat{p_D}(a|d)$ through Monte Carlo simulation.
  - Assessment of $P_A(\cdot)$ typically based on $p_D(\cdot)$
    - Dirichlet distribution (process) for discrete (continuous).
  - For U_A, information about Attacker's interests
    - Aggregate with weighted measurable value function.
    - Assume risk proneness.
    - Distributions over weights and risk proneness coefficient.

3. A framework for risk analysis and adversarial risk analysis

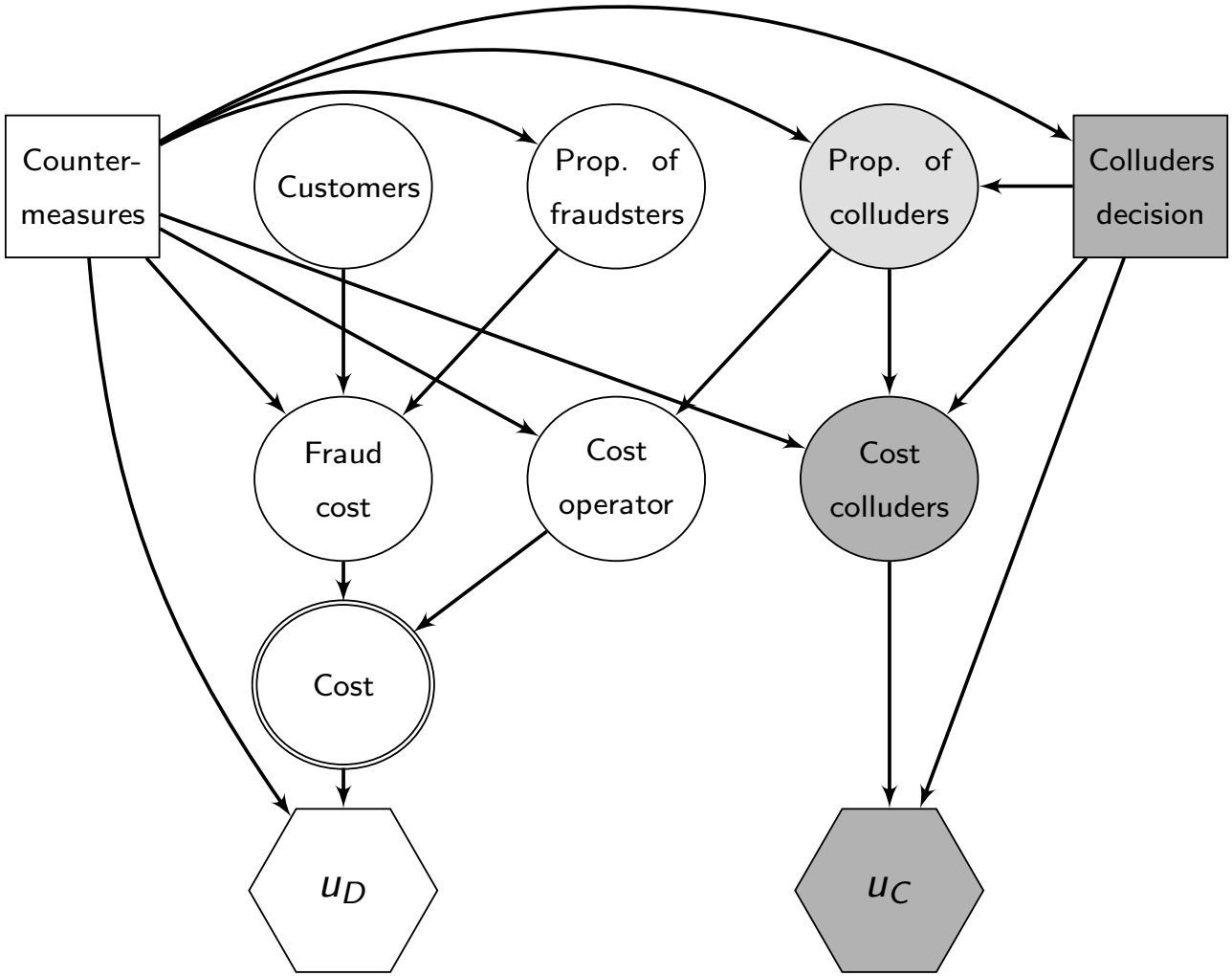
General influence diagram



- Incorporate uncertainty from non-adversarial threats.
- Defender's u_D aggregates consequences from both problems.

4. Case study: fighting fare evasion

Influence diagram



Description of problem

SECONOMICS

- Metro operator D protecting from:
  - Fare evasion. Two types of evaders:
    - Standard (standard random process).
    - Colluders (ARA; explicitly modeling intentionality).

		Role	Features
d_1	Inspector	Prev./rec.	Inspect customers. Collect fines
d_2	Door guard	Prev.	Control access points
d_3	Guard	Prev.	Patrol along the facility
d_4	Door	Prev.	New secured automatic access doors
d_5	Ticket clerk	Prev.	Current little implication in security

Feasible portfolios

SECONOMICS

- Associated unit costs q_1, q_2, q_3, q_4.
- $d_5 \in \{0,1\}$ ($d_5 = 1 \rightarrow$ clerks involved, incurred costs q_5).

$$q_1 d_1 + q_2 d_2 + q_3 d_3 + q_4 d_4 \leq B,$$
$$d_1, d_2, d_3, d_4 \geq 0,$$
$$d_1, d_2, d_3, d_4 \text{ integer},$$
$$d_4 \leq \bar{d}_4,$$
$$d_5 \in \{0,1\},$$

($\bar{d}_4$ maximum # of doors that may be replaced).

Defender's problem

- **Operator** invests $d = (d_1, d_2, d_3, d_4, d_5)$. (Constraints)
  - Fare evasion costs (partly mitigated by fines).
- $\phi(d)$ evaders proportion. $q(d_1)$ inspection proportion

$$\phi(d) = \phi_0 \cdot \exp\left(-\sum_{k=1}^{5} \gamma_k d_k\right) + \phi_r.$$

  - γ_k's $\rightarrow$ effect of $(d_1, d_2, d_3, d_4, d_5)$ on fraud proportion.
  - $(\phi_0 + \phi_r)$ current fraud proportion.
  - ϕ_r residual proportion even with infinite resources.
  - Each new inspector $\rightarrow$ # inspected tickets (nonlinear increase).
- $1 - \phi(d) \longrightarrow N_1$ civic customers pay ticket.
- $\phi(d)[1 - q(d_1)] \longrightarrow N_2$ not pay, not caught (loss v).
- $\phi(d)q(d_1) \longrightarrow N_3$ do not pay but caught (income f).

Attacker's problem

- **Colluders** see security investments d (Seq D-A).
- Fare evasion proportion $r \to r'$, inspection proportion $q_A(d_1)$
  - $1 - r' \to M_1$ pay, abortion (income v).
  - $r'(1 - q_A(d_1)) \to M_2$ not pay, not caught (loss v).
  - $r' q_A(d_1) \to M_3$ not pay, caught (income f).
- Operational costs, including preparation costs q

$$c_A = v(M_2 - M_1) - fM_3 - rqM.$$

- Colluders risk prone in benefits

$$u_A(c_A) = \exp(k_A \cdot c_A),\ k_A > 0.$$

- **Target**: Assess $h(r|d)$, Defender's beliefs over proportion of evasion attempts given d.

Solving the Defender's problem

- Operator benefit/cost balance

$$c_D(N_1, N_2, N_3, M_1, M_2, M_3, d) = - v(N_2 + M_2) + f(N_3 + M_3) - \sum_{k=1}^{5} q_k d_k.$$

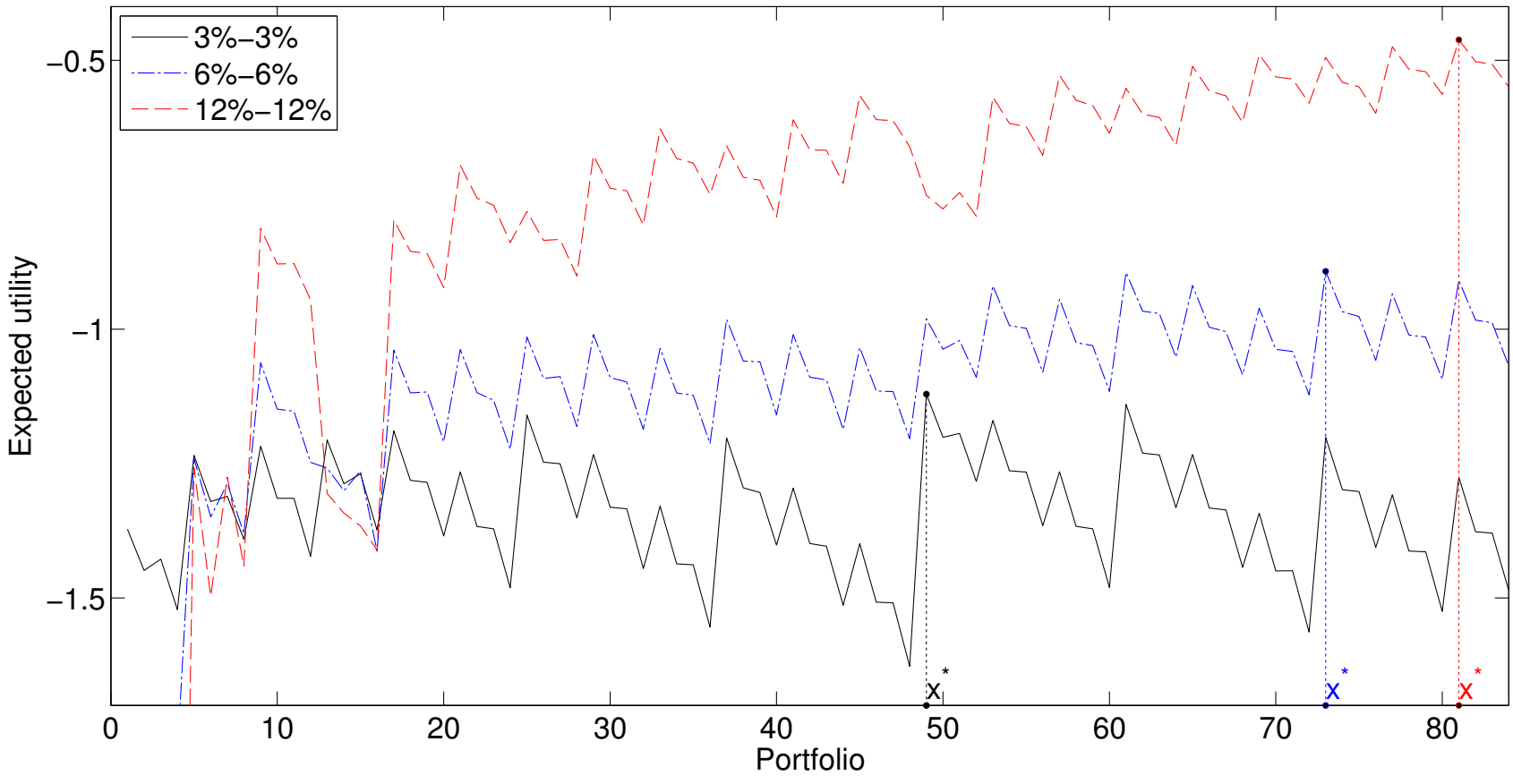
- Operator risk averse to increase in income,

$$u_D(c_D) = -\exp(-k_D \cdot c_D).$$

- Evaluate security plan d maximizing expected utility

$$\psi_D(x) = \int \left[\sum_{\substack{N_1, N_2, N_3 \\ M_1, M_2, M_3}} p_{M_1 M_2 M_3 d} \times p^1_{N_1 d}\, p^2_{N_2 d}\, p^3_{N_3 d} \times u_D(c_D) \right] \times h(r|d)\, \mathrm{d}r.$$

Results


3%–3%
6%–6%
12%–12%
Expected utility
−0.5
−1
−1.5
0
10
20
30
40
50
60
70
80
Portfolio
x*
x*
x*

Results

$p_0+p_r=0.03, M=30000$				$p_0+p_r=0.06, M=60000$				$p_0+p_r=0.12, M=120000$			
x	Invest.	$\psi(x)$	Income	x	Invest.	$\psi(x)$		x	Invest.	$\psi(x)$	
$(1,0,0,0,0)$	50000	-1.12	-22826	$(1,2,0,0,0)$	100000	-0.89		$(1,3,0,0,0)$	125000	-0.46	
$(1,0,0,0,0)$	50000	-1.12	-22826	$(1,0,0,0,0)$	50000	-0.98		$(1,0,0,0,0)$	50000	-0.75	
$(0,3,0,0,0)$	75000	-1.20	-36797	$(0,3,0,0,0)$	75000	-0.98		$(0,3,0,0,0)$	75000	-0.66	
$(0,0,2,0,0)$	60000	-1.22	-39409	$(0,0,2,0,0)$	60000	-1.06		$(0,0,2,0,0)$	60000	-0.81	
$(0,0,0,1,0)$	15000	-1.43	-71255	$(0,0,0,1,0)$	15000	-1.82		$(0,0,0,1,0)$	15000	-3.52	
$(0,0,0,0,1)$	15000	-1.45	-74147	$(0,0,0,0,1)$	15000	-2.10		$(0,0,0,0,1)$	15000	-4.19	
$(0,3,2,1,1)$	150000	-1.63	-97348	$(0,3,2,1,1)$	150000	-1.20		$(0,3,2,1,1)$	150000	-0.66	
$(0,3,2,1,0)$	150000	-1.51	-82303	$(0,3,2,1,0)$	150000	-1.11		$(0,3,2,1,0)$	150000	-0.61	

- Optimal portfolio $d^*=(1,0,0,0,0)$, with $\psi(x)=-1.12$, associated investment 50,000 euros, and expected losses 22,826 euros (investment plus expected balance between fraud and collected fines, $+27,174$ euros).
- Results sensitive to variations in evasion proportion $\phi_r+\phi_0$.
  - Operator needs higher investments for higher proportions.
  - Essential that inspectors really carry out their task.

Conclusions

SECONOMICS

- RA+ARA methodology.
- Sequential Defend-Attack model as basic template.
- Expand basic template with additional uncertainty nodes.
- Case study in metro security $\rightarrow$ fare evasion.

Current methodological developments

SECONOMICS

- ARA (Ríos Insua et al., 2009) approach for multithreat problem over multiple sites. (Ríos Insua et al., 2014b)
  - Multiple uncoordinated attacks.
    - Outcome of attacks might affect each other.
  - Extension to multiple sites.
    - Sequential Defend-Attack for each site/threat.
    - Models related by resource constraints and value aggregation.
    - No particular spatial structure.
  - Case study: metro network protection against
    - Fare evasion. (Ríos Insua et al., 2014a)
    - Pickpocketing by a team.

Future developments

- Multiple defenders and eventual coordination.
- Coordination and rationality type of attacks.
- More complex interactions between defenders and attackers.
- Mobility of resources.

Acknowledgments

- This project has received funding from the **European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under grant agreement no 285223**.
- Work has been also supported by the Spanish Ministry of Economy and Innovation program MTM2011-28983-C03-01, the Government of Madrid RIESGOS-CM program S2009/ESP-1685 and the **AXA-ICMAT Chair on Adversarial Risk Analysis**.
- Grateful to TMB experts and stakeholders for fruitful discussion about modeling issues.

Bibliography

- Ríos Insua, D., J. Cano, M. Pellot, R. Ortega. 2014. *Current Trends in Bayesian Methodology with Applications*, chap. From Risk Analysis to Adversarial Risk Analysis. CRC Press, To appear.
- Ríos Insua, D., J. Cano, M. Pellot, R. Ortega. 2014. Multithreat Multisite Protection: A Case Study in Metro Security. *Submitted for publication*.
- Ríos Insua, D., J. Ríos, D. Banks. 2009. Adversarial risk analysis. *Journal of the American Statistical Association* 104(486) 841–854.