Assessing threats by reasoning about the adversary

- Threat assessment is commonly addressed as a pattern recognition task using predefined templates of threatening situations...
- ... but assessing threats is largely about predicting unprecedented and unexpected behaviors
- This requires the ability to reason about the adversary's intentions so as to predict his behavior, anticipate his actions, and be proactive in deploying effective countermeasures

⇒ Adversarial Plan Recognition

A generative game-theoretic framework

- Why generative?
  ⇒ no need for a predefined plan library
- Why game-theoretic?
  ⇒ a natural and popular formalism for representing strategic interactions between competitive agents
- The library of possible (optimal) plans for a rational adversary is obtained by solving a set of Adversarial Plan Recognition Stochastic Games (APRSGs)
- This set of APRSGs is generated automatically from the definition of a planning problem, using a contextual action model

PRESAGE : Adversarial Plan Recognition using Stochastic Games

Plan Recognition: APRSGs generation and solving

Optimal plan library

Plan Recognition

High-level functional architecture

Converting domain knowledge to stochastic games

Planning problem definition:

\[ P = (Ag, I, G, T) \]

with \( Ag \) a finite set of actors, \( I \) the initial situation, \( G \) the set of possible goals of the adversary, \( T \) a set of terminal situations

Assessing opportunities for action

- \( O_s \) is a mapping that assigns to each non-terminal situation \( s \) the set of joint opportunities available in \( s \)
- Opportunity \( o' = (o, t, p) \) with \( o \) an action, \( t \) the target of the action, and \( p \) the probability of success of action \( o \) when performed on \( t \)

SITUATION PROJECTION

- Given planning problem \( P \), we generate the goal-independent state space \( S^T \), and the transition function \( T \)

Stochastic games generation

- Given the projected situation and a set of payoff functions \( \{R_{o,i} : o \in Ag, g \in G\} \),
  we generate one APRSG for each goal \( g \in G \)

Inferring the plan of a rational adversary

- The set \( \Pi_P = \{\pi^g_i | g \in G\} \) of optimal policies for the adversary and the set \( \beta_P(\Pi_P) \) of best responses given problem \( P \) is obtained by computing one Nash equilibrium for each \( \Gamma_g \)
- Let \( o_i \) be the opportunity played by the adversary in state \( s_i \). Given history \( H_i(t) \), the belief that the adversary is following policy \( \pi^g_i \in \Pi_P \) at \( t \) is given by:

\[ b_{i}^{\pi^g_i}(o_i|\pi^g_i, s_i) \propto \prod_{j=1}^{t-1} P(o_{j-1}|\pi^g_i, s_{j-1}) T_{j}(s_{j-1}, o_{j-1}, s_{j+1}) \]

Ongoing and future works

- Current limitations - Full observability, fully rational adversary, scalability issues, application-dependent payoff function, two-player and zero-sum constraints
- Future works - Adversary with bounded rationality, actively hostile to the recognition process (use of deception and concealment) → partially observable situation, generic payoff function