Game theoretical approaches to actionable cyber-attack forecasts

João P. Hespanha & Jeff Shamma
Adversarial forecasting

Decision making in non-adversarial environments:
1. Given
   • noisy observations (e.g., data streams from sensors)
   • uncertain model (e.g., expressing possible attack progression)
2. Determine
   • estimate of the state of the system (e.g., attack source)
   • predict future outcomes (e.g., probability of mission success)

Key assumption:
Noise in observations & model uncertainty are independent of our actions

Decision making in adversarial environments:
1. & 2. same, but…
   Noise in observations & model uncertainty are likely to conspire against us – Controlled by a reactive rational adversary

   *Events that would be unlikely in a stochastic setting may become very likely in an adversarial setting.*

Game theory provides a framework to reason about adversarial environments
Game theoretical formulation

Framework:
1. Dynamic models for the missions
   • causal relations between potential actions and ultimate mission success:
     “if action X is taken, then outcome Y is expected”
   • probabilistic characterizations can be used to model uncertainty
   • actions of potentially reactive adversary taken into account
   • models take form of Competitive Markov Decision Processes (CMDP)
2. Forecasting corresponds to estimating the probability of reaching specific states (ultimately, mission success)
   • taking into account actions of reactive adversary & countermeasures by network security system
Cyber-therapy domain poses challenges to mainstream game theory

1. Partial information
   security system (and adversaries) do not possess perfect information about system’s state (e.g., has a particular system been compromised)

2. Uncertain action spaces
   security system may not know all possible adversary actions (which attacks can the adversary attempt?)

3. Uncertain adversary intent
   security system may not know the adversary’s ultimate objective (e.g., is the ultimate goal to prevent the success of the basic mission under consideration, or just create general disruption?)

uncertainty, uncertainty, uncertainty => complexity

Solution needs to be robust with respect to sensor errors & to errors in modeling adversary behavior
Key novel concepts:

• considering affect of uncertainty in strategy construction – robustness with respect to novel and unexpected threats

• use of roll-out strategies – estimate probability of mission success taking into account potential actions of adversary and counter measures

• use of adversarial detection – prevent manipulation by an adaptive adversary
Uncertainty management

What actions are available to adversary?
What information is available to adversary?
What is the specific intent of the adversary?

Uncertainty tradeoffs
• player must balance between
  1. acting towards asset protection
  2. probing adversary’s capabilities and intent
• player must balance between
  1. exploiting available information
  2. revealing information to the adversary

Two approaches:
• learning-based approaches find solutions by observing outcomes (without assumptions on adversary models)
• model-based approaches can exploit deception in partial information games (care must be taken to create robustness to modeling errors)
Not feasible to compute optimal (Nash equilibrium) strategies
• computational complexity
• optimality may lead to fragility with respect to modeling errors

Roll-out strategies (approximate dynamic programming)
• simulation-based approach
• optimize near-term actions based on
  1. simulation of possible evolutions (short term)
  2. baseline long-term strategies (suboptimal)
• simulation will consider
  • (stochastic) uncertainty
  • adversarial actions
• potentially hierarchical (baseline strategies may be roll-out strategies)
Roll-out strategies (approximate dynamic programming)

- simulation-based approach
- optimize near-term actions based on
  1. simulation of possible evolutions (short term)
  2. baseline long-term strategies (suboptimal)
- simulation will consider
  - (stochastic) uncertainty
  - adversarial actions
- potentially hierarchical (baseline strategies may be roll-out strategies)

Has been successfully used in the past, but for
1. Optimization (stochasticity, but nonadversarial)
2. Deterministic, highly structured games
3. Simple stochastic games
Detection in adversarial environments

Traditional hypothesis testing is prone to manipulation by adversaries
• introduce randomization into attacks to hide patterns
• learn detection thresholds

Adversarial detection
• detection system wants to minimize probability of false positive/negatives; adversary want to maximize such probabilities
• viewed as a repeated-game: two players continuously adjust their strategies
• no-regret methods
  • performance-based approach
  • strategies evolve based on retrospective analysis (what would have been the optimal decision…)
  • converge to no-regret solutions
  • do not require explicit knowledge of adversaries ultimate goal and abilities
Detection in adversarial environments

Traditional hypothesis testing is prone to manipulation by adversaries
- introduce randomization into attacks to hide patterns
- learn detection thresholds

Adversarial detection
- can minimize learning time by introducing dynamic adaptation (respond not just to what the adversary is doing now, but also on inferred trends)
- can make system more responsive by introducing presumed models for the adversary
  - must consider uncertainty in adversary model to avoid fragilities
  - adversary model needs to be reactive and dynamic (not just a scripted adversary)
Adversary aware forecasts will…

• increase robustness with respect to novel and unexpected threats

• permit the estimation of the probability of mission success taking into account potential actions of adversary and counter measures

• prevent manipulation by an adaptive adversaries