SAND2011-4697 C

SAND2011-4697C

Proactive Defense for Evolving Cyber Threats

Rich Colbaugh*† **Kristin Glass**†

*Sandia National Laboratories
†New Mexico Institute of Mining and Technology

July 2011

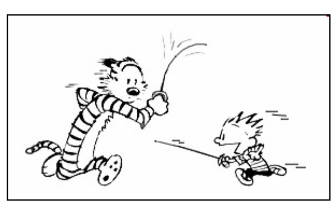
Introduction

Objective

Characterize predictability of the cyber attacker/defender "arms race" and leverage findings to create a framework for designing *proactive* defenses for large computer networks.

Outline

- Adversarial dynamics: predictability of non-transitive games.
- Responsive defense:
  transfer learning, sample results.
- Proactive defense:
  synthetic attack generation, sample results.



Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Adversarial Dynamics

Adversarial data mining

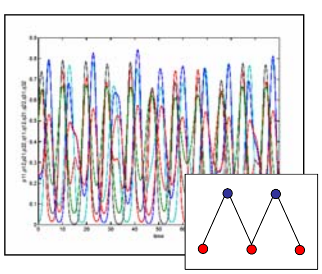
- Coevolutionary adversarial dynamics are central in a broad range of important phenomena, including
  - security-related (e.g., terrorism, cyber defense, border security, proliferation);
  - business-related (e.g., marketing, economics, finance, fraud).

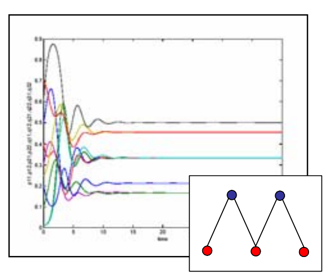
  However, "data mining" algorithms typically assume that the data-generating process is independent of the algorithm's activities.
- We conjecture that coevolution of adversary strategies generates dynamical structures which can be exploited to design proactive defenses that are effective against both current and near future attacks.

Adversarial Dynamics

Predictability of adversarial coevolution

- Influential work by [Farmer et al. 2002] suggests that, for non-transitive games (e.g. rock-paper-scissors), *reactive* adversarial learning results in unpredictable dynamics.
- Our work shows broad classes of *proactive* learning leads to predictable dynamics and suggests utility of extrapolating adversary behavior into the near future.





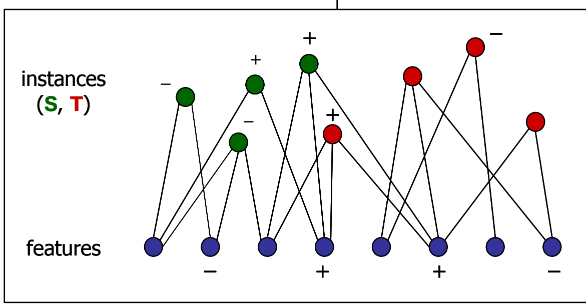
Responsive Defense

Problem

Increase responsiveness of network defenses by exploiting attacker-defender coevolution via bipartite graph-based transfer learning.

Approach

bipartite graph data model



$$\min_{c_{aug}} \; c_{aug}^{T} L c_{aug} + \beta_1 \left\| d_{S,est} - k_S d_S \right\|^2 + \beta_2 \left\| d_{T,est} - k_T d_T \right\|^2 + \beta_3 \left\| c - w \right\|^2$$

objective function for learning

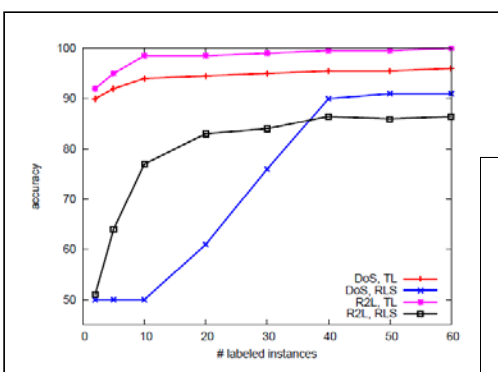
Responsive Defense

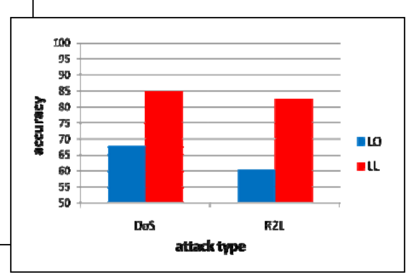
Sample results

Intrusion detection with (publicly-available) KDD Cup 99 dataset.


UCI KDD Archive

Welcome to the UCI Knowledge Discovery in Databases Archive





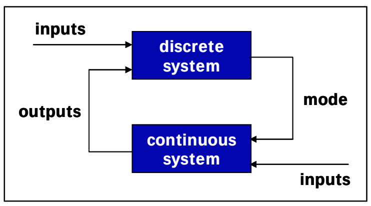
Proactive Defense

Problem

Enable proactive network defense by generating "predicted" attack data and using this synthetic data to train defense systems.

Approach

S-HDS model



Synthetic Data Learning Algorithm

1. Identify relevant modes of attack (e.g., via SMEs or auxiliary data).
2. Construct S-HDS model and generate set of synthetic attack instances A_S.
3. Assemble sets of normal network activity N and measured attack activity A_M for network of interest.
4. Train classifier (e.g., RLS) using training data $TR = N_M \cup A_M \cup A_S$. Estimate class label (innocent or malicious) of any network activity x with formula: $orient(x) = sign(c^T x)$.

Proactive Defense

Sample results

- Setup: attacker (Spammer) assumes defender (Spam filter) uses naïve Bayes (NB) for detection and manipulates observable (email message) to defeat NB.
- Proactive Spam filter design:
  - generate *synthetic Spam* data via Algorithm SDL with two attack modes (add-words, synonyms);
  - train proactive filter on both real current Spam and synthetic (near future) Spam;
  - results shown are for Ling-Spam dataset.

NB Algorithm: Nominal Spam

class\truth	non-Spam	Spam
non-Spam	262	19
Spam	1	215

NB Algorithm: Nominal and Attack Spam

class\truth	non-Spam	Spam
non-Spam	524	253
Spam	2	215

Algorithm SDL: Nominal and Attack Spam

class\truth	non-Spam	Spam
non-Spam	524	40
Spam	2	428