

# 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.



## Adversarial Dynamics

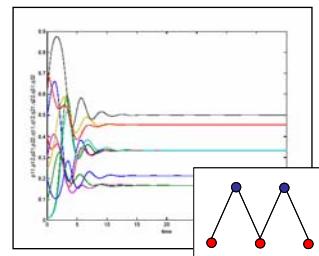
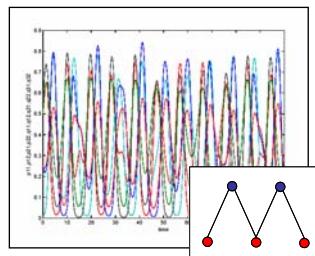
### Adversarial data mining

- 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**

**objective function for learning**

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

# 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

Administrator's note (July 21, 2009): We no longer maintain this web page as we have merged the KDD archive with the UCI Machine Learning Archive. For any questions, please contact us at [mlarchive@ics.uci.edu](mailto:mlarchive@ics.uci.edu).

This is a smaller repository of large data sets which encompasses a wider variety of data types, machine tasks, and application areas. The primary role of this repository is to enable researchers in knowledge discovery and data mining to solve interesting and better data analysis algorithms to very large and complex data sets.

Creation of this archive was supported by a grant from the Information and Data Management Program at the National Science Foundation. This archive is part of the UCI Machine Learning Repository, which contains many more accessible data sets for research in KDD and data mining. It complements the original UCI Machine Learning Archive, which typically focuses on smaller classification-oriented data sets.

In addition to storing data and description files, we also archive task files that describe a specific analysis, such as clustering or regression, for the data sets stored. The [mlarchive.ics.uci.edu](http://mlarchive.ics.uci.edu) lists typical data types and tasks of datasets.

# labeled instances	Dos, TL	Dos, RLS	R2L, TL	R2L, RLS
0	90	50	90	50
10	92	52	98	65
20	93	60	98	80
30	94	82	98	85
40	95	88	98	88
50	96	90	98	88
60	96	90	98	88

attack type	LO	LL
Dos	68	84
R2L	58	82

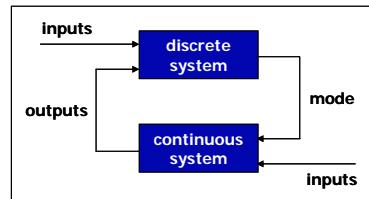
## Proactive Defense

### 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:  $\text{orient}(x) = \text{sign}(c^T x)$ .

### 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