SAND2012-4007C

Predictive Moving Target Defense

Richard Colbaugh
Sandia National Laboratories
Albuquerque, NM USA
colbaugh@comcast.net

Kristin Glass
New Mexico Institute of Mining and Technology
Socorro, NM USA
kglass@icasa.nmt.edu

***Abstract*—There are substantial potential benefits to considering *predictability* when designing defenses against adaptive adversaries, including the possibility to develop defense systems capable of anticipating adversary behavior while reducing the adversaries' ability to predict defensive actions. This paper adopts such a perspective, leveraging the coevolutionary relationship between attackers and defenders to derive methods for predicting and countering adversary attacks and for limiting the extent to which adversaries can learn about the defense strategies. The proposed approach combines game theory with machine learning to model adversary adaptation in the learner's feature space, thereby producing classes of predictive and "moving target" defenses which are scientifically-grounded and applicable to problems of real-world scale and complexity. Case studies with a large cyber security dataset assembled for this investigation demonstrate that the proposed algorithms outperform gold-standard techniques, offering effective and robust defense against evolving adversaries.**

***Keywords*—**predictive defense, moving target defense, game theory, machine learning, adaptive adversaries, cyber security.

I. INTRODUCTION

Adaptive adversaries are a principal concern in many security domains, including cyber defense, border security, counterterrorism, and crime prevention [e.g. 1-3]. Consequently, there is great interest in developing defenses which maintain their effectiveness despite evolving adversary strategies and tactics. A potentially powerful approach to pursuing such goals is to explicitly consider system *predictability,* for instance in order to design defenses which are able to anticipate adversary behavior and/or decrease their own predictability. Preliminary studies that employ predictability assessment in a cyber security context include [4,5].

It is natural to formulate this defense design problem as a game [6], in which the defense attempts to predict and counter adversary actions while reducing its own predictability. Unfortunately, previous game-theoretic approaches to adversary defense [e.g. 7-12] have encountered numerous challenges, and we mention two that have been especially daunting. First, the set of possible attacker actions is typically very large in real-world settings, and because the complexity of most game models increases exponentially with the number of actions available to the players, this has often made these models intractable in practice. Second, it has proved difficult to derive models that capture evolving attacker behavior in any but the most idealized situations.

In this paper we overcome these challenges by developing game-based models for attack-defend interaction within a machine learning (ML) framework [13], enabling design of robust defenses for practical applications. We formulate the defense task as one of behavior classification, in which innocent and malicious activities are to be distinguished, and assume only limited information is available regarding prior attacker behavior or attack attributes. The defense classifiers model attacker actions in ML *feature space,* that is, in the space of variables the ML algorithms use for learning. Formulating attack prediction/defense synthesis in this "compressed" and abstract space enables the derivation of algorithms that can be applied to practical, real-world problems.

The first of the proposed defense systems attempts to predict and counter adversary adaptation as a means of providing effective defense against both current and future attacks. A key step in the approach is modeling the way attackers *adapt* their behaviors rather than modeling the behaviors themselves. Crucially, the proposed approach seeks to design optimal defenses for evolving attacks, rather than to predict new attacks perfectly, and therefore enjoys robust performance in the presence of (inevitable) prediction errors. To permit the performance of this predictive defense method to be evaluated, we have assembled a large collection of Spam and non-Spam emails reflecting the evolution of Spammer tactics over a nine year period. A case study with this dataset demonstrates that the proposed defense outperforms a gold-standard Spam filter.

An important consideration when applying classifier-based defense techniques, even predictive ones, is the extent to which adversaries can reverse-engineer the learning algorithm and use this knowledge to circumvent the defense. The goal of the second proposed defense is thus to reduce defense system predictability and increase the difficulty of the adversary's reverse-engineering task. We adopt a "moving target" (MT) perspective, in which the defense presents a dynamic posture to the adversaries as a way of increasing the adversaries' uncertainty concerning defense operation [14]. By leveraging recent advances in the theory of repeated, incomplete information games [15], we derive a simple MT defense procedure and prove its optimality; interestingly, the optimal MT schedule can be specified independently of the details of the adversaries' strategies. The efficacy of the proposed MT defense is evaluated via a case study with the set of Spam and non-Spam emails mentioned above. This test reveals that the MT defense substantially outperforms a well-tuned static Spam filter against adaptive adversaries.

II. PREDICTIVE DEFENSE

A. Problem Formulation

There are significant potential benefits to developing *predictive* methods of defending against adaptive adversaries, in which opponents' evolving strategies are anticipated and these insights are employed to counter novel attacks. This section considers the following concrete instantiation of the predictive defense problem: given some history of attacker actions, design a defense system which performs well against both current and future attacks. It is reasonable to expect that concepts and techniques from game theory might be helpful in understanding adversary adaptation, and indeed such approaches have been explored in a variety of domains [e.g. 7-12]. However, as indicated in the Introduction, these investigations have encountered scalability and complexity challenges which have limited their practical utility. In this section we address these challenges by deriving our game-based model within an ML framework, enabling effective defense in realistic settings. (See [16] for a general discussion of the value of combining behavioral modeling with data mining algorithms for discovery and prediction applications.)

We approach the task of countering adversarial behavior as an ML classification problem, in which the objective is to distinguish innocent and malicious activity. Each instance of activity is represented as a feature vector $x \in \Re^{|F|}$, where entry x_i of x is the value of feature i for this instance and F is the set of instance features. In what follows, F is a set of "reduced" features, obtained by projecting measured feature vectors into a lower-dimensional space. While feature reduction is standard practice in ML [13], we show below that *aggressive* reduction allows us to efficiently manage the complexity of our game models. Behavior instances x belong to one of two classes: positive/malicious and negative/innocent (generalizing to more than two behavior classes is straightforward [13]). The goal is to learn a vector $w \in \Re^{|F|}$ such that classifier orient = $\text{sign}(w^T x)$ accurately estimates the class of behavior x, returning +1 (–1) for malicious (innocent) activity.

As indicated above, it is useful to assess the predictability of a phenomenon before attempting to predict its evolution; for example, such an analysis permits identification of measurables that possess predictive power [17]. There has been limited theoretical work assessing predictability of adversarial dynamics, but existing studies suggest attack-defend coevolution often generates predictable dynamics. For instance, although [18] finds that certain player strategies lead to chaos in a simple repeated game, [16] shows that large sets of player strategies and repeated games exhibit predictable adversarial dynamics. Here we supplement this theoretical work by conducting an empirical investigation of predictability, and select as our case study a cyber security problem – Spam filtering – which possesses attributes that are representative of many adversarial domains.

To conduct this investigation, we first obtained a large collection of emails from various publicly-available sources for the period 1999-2006, and added to this corpus a set of Spam emails acquired from B. Guenter's Spam trap for the same time period. Following standard practice, each email is modeled as a "bag of words" feature vector $x \in \Re^{|F|}$, where the entries of x are the frequencies with which the words in vocabulary F appear in the message. The resulting dataset consists of ~128,000 emails composed of more than 250,000 features. We extracted from this collection of Spam and non-Spam emails the set of messages sent during the 30 month period between January 2001 and July 2003 (other periods exhibit very similar behavior). Finally, the dimension of the email feature space was reduced via singular value decomposition (SVD) analysis [13], yielding a reduction in feature space dimension of four orders of magnitude (from ~250K to 20).

We wish to examine, in a simple but meaningful way, the predictability of Spam adaptation, and propose two intuitively reasonable criteria with which to empirically evaluate predictability: *sensibility* and *regularity* (a comprehensive theoretical framework for defining and assessing predictability is given in [17]). More specifically, and in the context of Spam, it would be *sensible* for Spammers to adapt their messages over time in such a way that Spam feature vectors x_S comes to resemble the feature vectors x_{NS} of legitimate emails, and *regularity* in this adaptation might imply that the values of the individual elements of x_S approach those of x_{NS} in a fairly monotonic way.

To permit convenient examination of the evolution of feature vectors x_S and x_{NS} during the 30 month period under study, the emails were first binned by quarter. Next, the average values for each of the 20 (reduced) features was computed for all the Spam emails and all the non-Spam emails (separately) for each quarter. Figure 1 illustrates the feature space dynamics of Spam and non-Spam messages for one representative element (F1) of this reduced feature space. As seen in the plot, the value of feature F1 for Spam approaches the value of this feature for non-Spam, and this increasing similarity is a consequence of changes in the composition of Spam messages (the value of F1 for non-Spam emails is essentially constant). The dynamics of the other feature values (not shown) are analogous.

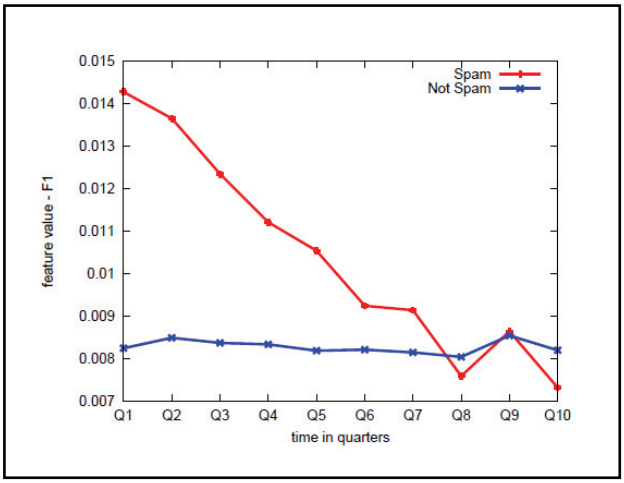

Figure 1. Spam/non-Spam evolution in feature space. The plot depicts evolution of feature F1 for Spam (red) and non-Spam (blue) over time (horizontal axis).

Observe that the Spam dynamics illustrated in Figure 1 reflect *sensible* adaptation on the part of Spammers: the features of Spam email messages evolve to appear more like those of non-Spam email, making Spam more difficult to detect. Additionally, this evolution is *regular*, with feature values for Spam approaching those for non-Spam in a nearly-monotonic fashion. Thus this empirical analysis indicates that coevolving Spammer-Spam filter dynamics possesses some degree of predictability, and that the features employed in Spam analysis may have predictive power; this result is in general agreement with the conclusions of the theoretical predictability analysis reported in [16]. Moreover, because many of the characteristics of Spam-Spam defense coevolution are shared by other adversarial systems, this result suggests these other systems may have exploitable levels of predictability as well.

B. Predictive Defense Algorithm

The proposed approach to designing a predictive defense system which works well against both current and future attacks is to combine ML with a simple game-based model for adversary behavior. In order to apply game-theoretic methods, it is necessary to overcome the complexity and model-realism challenges mentioned above. We address problem complexity by modeling adversary actions directly in an aggressively-reduced ML feature space, so that the (effective) space of possible adversary actions which must be considered is dramatically decreased. The difficulty of deriving realistic representations for attacker behavior is overcome by recognizing that the actions of attackers can be modeled as attempts to *transform* data (i.e., feature vectors x) in such a way that malicious and innocent activities are indistinguishable. (This is in contrast to trying to model the attack instances "from scratch"). It is possible to model attacker actions as transformations of data because, within an ML problem formulation, historical attack data are available in the form of training instances.

We model adversarial coevolution as a sequential game, in which the attacker and defender iteratively optimize the following objective function:

$$\min_{w} \max_{a} \left[-\alpha\|a\|^3 + \beta\|w\|^3 + \sum_i \text{loss}\left(y_i, w^T(x_i + a)\right) \right] \qquad (1)$$

In (1), the loss function represents the misclassification rate for the defense system, where $\{y_i, x_i\}_{i=1}^n$ denotes pairs of "nominal" activity instances x_i and labels y_i, and vector w parameterizes the defense (recall that the defense attempts to distinguish malicious and innocent activity using the classifier orient = sign($w^T x$)). The attacker attempts to circumvent the defense by transforming the data through vector $a \in \Re^{|F|}$, and the defender's goal is to counter this attack by appropriately specifying classifier vector $w \in \Re^{|F|}$. The terms $-\alpha\|a\|^3$ and $\beta\|w\|^3$ define "regularizations" imposed on attacker and defender actions, respectively, as discussed below.

Observe that (1) models the attacker as acting to increase the misclassification rate with vector a, subject to the need to limit the magnitude of this vector (large a is penalized via the term $-\alpha\|a\|^3$). This model thus captures in a simple way the fact that the actions of the attacker are in reality always constrained by the goals of the attack. For instance, in the case of Spam email attacks, the Spammer tries to manipulate message x in such a way that it "looks like" legitimate email and evades the Spam filter w. However, transformed message x+a must still communicate the desired information to the recipient or the attacker's goal will not be realized, and so the transformation vector a cannot be chosen arbitrarily.

The defender attempts to reduce the misclassification rate with an optimal choice for vector w, and avoids "over-fitting" through regularization with the $\beta\|w\|^3$ term [13]. Notice that the formulation (1) permits the attacker's goal to be modeled as counter to, but not exactly the opposite of, the defender's goal, and this is consistent with many real-world settings. Returning to the Spam example, the Spammer's objective of delivering messages which induce profitable user responses is not the inverse of an email service provider's goal of achieving high Spam recognition with a very low false-positive rate.

The preceding development can be summarized by stating the following predictive defense (PD) algorithm:

Algorithm PD

1. Collect historical data $\{y_i, x_i\}_{i=1}^n$ which reflects past behavior of the attacker and past legitimate behavior.
2. Optimize objective function (1) to obtain the predicted actions a* of the attacker and the optimal defense w* to counter this attack.
3. Estimate the status of any new activity x as either malicious (+1) or innocent (−1) via orient = sign($x^T w^*$).

Observe that Step 2 of this algorithm can be interpreted as first predicting the attacker strategy through computation of attack vector a*, and then learning an appropriate countermeasure w* by applying ML to the "transformed" data $\{y_i, x_i + a^*\}_{i=1}^n$.

C. Algorithm Evaluation

This case study examines the performance of Algorithm PD for the Spam filtering problem. We use the Spam/non-Spam email dataset introduced above, consisting of ~128,000 messages that were sent during the period 1999-2006. The study compares the effectiveness of Algorithm PD, implemented as a Spam filter, with that of a well-tuned naïve Bayes (NB) Spam filter [5]. Because NB filters are widely used and work very well in Spam applications, this filter is referred to as the gold-standard algorithm. We extract from our dataset the 1000 oldest legitimate emails and 1000 oldest Spam messages for use in training both Algorithm PD and the gold-standard algorithm. The email messages sent during the four year period immediately following the date of the last training email are used as test data. More specifically, these emails are binned by quarter and then randomly sub-sampled to create balanced datasets of Spam and legitimate emails for each of the 16 quarters in the test period.

Recall that Algorithm PD employs aggressive feature space dimension reduction to manage the complexity of the game-based modeling process. This dimension reduction is accomplished here through SVD analysis, which reduces the dimension |F| of feature vectors from ~250K to 20) [13]. (The orthogonal basis used for this reduction is derived by performing SVD analysis using the 1000 non-Spam and 1000 Spam training emails.) Note that good classification accuracy can be ob-

tained with a wide range of (reduced) feature space dimensions. For example, a filtering accuracy of ~97% is achieved with the training data when using an NB classifier implemented with feature dimension ranging from |F|=100,000 to |F|=5.

The gold-standard strategy is applied as described in [5]. Algorithm PD is implemented with parameter values $\alpha = 0.001$ and $\beta = 0.1$, and with a sum-of-squares loss function. To evaluate the utility of the defenses against evolving adversaries, we train Algorithm PD and the gold-standard algorithm *once*, using the 1000 non-Spam/1000 Spam dataset, and then apply the filters without retraining to the four years of emails that follow these 2000 emails.

Sample results from this study are depicted in Figure 2. Each data point in the plots represents the average accuracy over ten trials (two-fold cross-validation). It can be seen that the filter based upon Algorithm PD significantly outperforms the gold-standard method: the predictive defense experiences almost no degradation in accuracy over the four years of the study, while the gold-standard method suffers a substantial drop in accuracy during this period. These results suggest that combining ML with simple game-based adversary models offers an effective means of defending against new attacks.

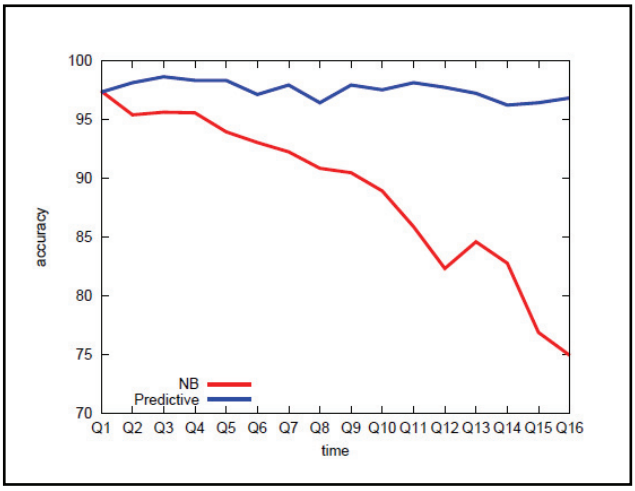

Figure 2. Results for predictive defense Spam filtering case study. The plot shows how filter accuracy (vertical axis) varies with time (horizontal axis) for the gold-standard NB filter (red) and Algorithm PD filter (blue).

III. MOVING TARGET DEFENSE

A. Problem Formulation

A defining characteristic of classification-based defense is the fact that adversaries continually attempt to reverse-engineer the classifier and use this knowledge to make informed adjustments to their behavior and circumvent the defense. One way to increase the difficulty of the adversary's reverse-engineering task is to employ moving target (MT) ideas, in which the defense adopts a time-varying posture in order to increase adversary uncertainty concerning defense operation [14]. In this section we derive an MT defense procedure which minimizes the predictability of defensive actions from the perspective of the adversary.

We conduct our development within the framework provided by two-player repeated games with incomplete information [15]. In these games one player, the *informed* player, has access to information that is unavailable to the other, *uninformed,* player. The informed player must weigh the relative benefits of exploiting her private information to achieve short-term advantage against the possibility that this exploitation may reveal information which results in the sacrifice of future gains. Because repeated incomplete information games explicitly account for the payoff-predictability tradeoff, they afford a convenient setting for deriving and comparing MT strategies.

Consider the following MT defense problem. Suppose the task of countering adversarial behavior is formulated as one of ML classification, in which the objective is to distinguish innocent and malicious activity. Each instance of activity is represented as a feature vector $x \in \Re^{|F|}$, where F is the set of ML features. Behavior instances x belong to one of two classes, positive/malicious and negative/innocent, and the goal is to learn a vector $w \in \Re^{|F|}$ such that classifier orient = sign($w^T x$) accurately estimates the class of behavior x.

A plausible way to reduce the degree to which adversaries can predict, and then adapt to and evade, the actions of a classifier is to introduce randomness into the way the ML features F are selected and used. One simple means of realizing this goal is delineated in the following three steps: 1.) divide the original feature set F into K randomly-selected, possibly overlapping subsets $\{F_1, \ldots, F_K\}$, where $|F_i| = m \; \forall \; i$; 2.) train one classifier for each feature subset F_i, yielding a collection of K classifiers $\{w_1, \ldots, w_K\}$; 3.) during operation, alternate between the classifiers w_i according to some random scheduling policy. In order to implement this MT defense, it is necessary to obtain a procedure for selecting which classifier is to be "active" at each time period. Thus the specific MT defense problem of interest to us can be stated: given a collection of classifiers $W=\{w_1, \ldots, w_K\}$, specify a policy for switching among the classifiers which minimizes defense predictability (from the point of view of the attacker).

B. Moving Target Scheduling Policy

A classifier schedule which minimizes defense predictability is sketched in the following theorem. Perhaps surprisingly, the optimal schedule is very simple to implement.

Theorem MT: Suppose we are given a collection of K classifiers $W = \{w_1, \ldots, w_K\}$ associated with randomly-selected feature subsets $\{F_1, \ldots, F_K\}$, an ecology of adversaries that wish to reverse-engineer the defense, and a sequence of times $t_1, t_2, \ldots$ at which it is permitted to switch classifiers. Defense system performance is optimized if, at each time t_i, the active classifier w_a is selected uniformly at random from the set W.

Proof (sketch): We model the interaction between an MT defense and an ecology of adversaries as a *hidden mode hybrid dynamical system* (HM-HDS); see, for instance, [15] for background on this class of dynamical systems. More specifically, the MT defense model is

$$\Sigma_{\text{HM-HDS}} = \{\text{CS}(w,a),\ W,\ P(w,a)\} \qquad (2)$$

where

- the *continuous system* CS(w,a) evolves according to some sequential attack-defend game dynamics (such as (1));
- the *discrete system* dynamics W, P(w,a) is a Markov chain with state set W (the set of candidate classifiers) and state transition probability matrix P(w,a), which in general may depend upon the continuous system state variables (w,a);
- the *hidden mode* is the currently active classifier $w_a \in W$.

A schematic of this HM-HDS model is depicted in Figure 3.

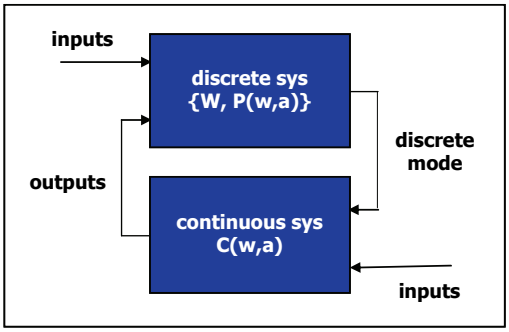

Figure 3. Schematic of basic HM-HDS feedback structure. The discrete and continuous systems in this framework model the selection of "active" classifier w_a and the resulting attack-defend dynamics, respectively.

The dynamics of the HM-HDS (2) evolve as follows. The discrete system specifies the currently active classifier w_a, and this information is communicated to the defender (but not the attacker) in the continuous system game. The attacker attempts to infer which classifier is active by observing defense actions, and computes attack vector a based on this estimate. The discrete system has access to the continuous system state (w,a) and may use this information in choosing the next active classifier.

We interpret these dynamics as a repeated incomplete information game, in which the discrete system is the informed player and the attacker-defender dynamics is the uninformed player. (This formulation, although less intuitive than the two-player game model adopted in Section II, facilitates analysis of MT dynamics.) The payoff to the discrete system is defined to be the negative of the misclassification rate, so that maximizing this payoff is equivalent to maximizing the defense classifier performance. This setup permits us to use the results in [15] to conclude that, provided the attacker has at least modest learning capabilities, the best strategy for the discrete system is to select the active classifier w_a uniformly at random from the set W at each time t_i. ■

Perhaps counterintuitively, the choice of new w_a does not depend upon the currently active classifier or the continuous state variables (w,a), because any such dependence has the potential to be exploited by the attacker.

C. *Algorithm Evaluation*

In this section we evaluate the effectiveness of the MT defense scheme summarized in Theorem MT. For convenience, we employ the same Spam filtering data and task introduced in Section II. To reduce complications in this initial assessment, a few simplifications are made:

- standard NB Spam filters are used for the classifiers $\{w_1, \ldots, w_K\}$ (rather than predictive filters generated by (1));
- only K=2 classifiers/feature subsets are used;
- attack vector a is computed in an optimal manner via (1), so that the adversary is assumed to be effective.

To enable the efficacy of the proposed MT defense to be quantified, its performance is compared to that of a well-tuned static NB filter trained using the full set of (reduced-dimension) features F. We examine a range of attack "strengths" by varying the parameter α in the optimization (1) (recall that the term $-\alpha\|a\|^3$ governs the magnitude of attack vector a). Attacks are normalized by assigning a unit attack strength AS (i.e., AS=1) to an attack with magnitude $\|a\|$ equal to the largest attack observed in the (real-world) Spam dataset.

We apply the static NB filter and the optimal two-mode (K=2) MT filter to the 2000 email training dataset described in Section IIC. Feature set F is taken to be the collection of 20 features with largest singular values, and feature subsets F_1 and F_2 are constructed by randomly sampling F (with replacement) until each subset contains 10 features. The filters are "attacked" by solving (1) for the optimal attack a* and then transforming Spam instances x according to the formula x+a*. To obtain a range of attack strengths, (1) is solved for different values of α, yielding the following AS values: AS=0, 0.25, 0.5, 0.75, 1.0, 1.25, 1.5 (so the attacks vary from 'no attack' to an attack magnitude 1.5 times larger than any seen in the Spam dataset).

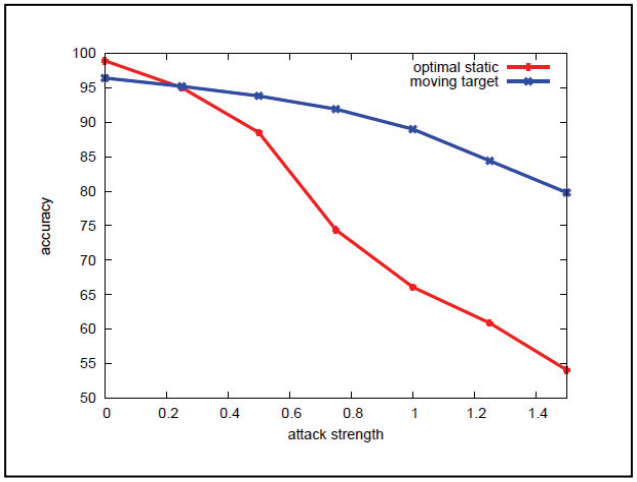

Figure 4. Results for moving target defense Spam filtering case study. The plot shows how filter accuracy (vertical axis) varies with attack strength (horizontal axis) for the static NB filter (red) and the optimally scheduled MT filter (blue).

Sample results are displayed in Figure 4. Each data point in the plots represents the average accuracy over ten trials (two-fold cross-validation). It can be seen that the filter based upon Theorem MT significantly outperforms the static NB filter. For instance, the MT defense achieves a classification accuracy of ~90% in the face of attacks of strength AS=1, compared with the ~65% accuracy obtained with the static filter, and provides an accuracy of ~80% for attacks with magnitude AS=1.5 while the static filter is only slightly more effective than random guessing (accuracy ≈ 54%) in this case. These results suggest that the proposed MT defense is capable of substantially increasing the difficulty of reverse-engineering tasks for even optimal attackers.

ACKNOWLEDGEMENTS

This work was supported by the Laboratory Directed Research and Development Program at Sandia National Laboratories. We thank Chip Willard and Curtis Johnson for numerous helpful discussions on aspects of this research.

REFERENCES

[1] *Proc. 2010 IEEE International Conference on Intelligence and Security Informatics*, Vancouver, BC Canada, May 2010.

[2] *Proc. 2011 IEEE International Conference on Intelligence and Security Informatics*, Beijing, China, July 2011.

[3] *Proc. 2012 IEEE International Conference on Intelligence and Security Informatics*, Washington, DC USA, June 2012.

[4] Colbaugh, R., "Does coevolution in malware adaptation enable predictive defense?", *IFA Workshop Series: Exploring Malware Adaptation Patterns*, San Francisco, CA, May 2010.

[5] Colbaugh, R. and K. Glass, "Proactive defense for evolving cyber threats", *Proc. 2011 IEEE ISI*, Beijing, China, July 2011.

[6] Peters, H., *Game Theory*, Springer, Berlin, 2008.

[7] Dalvi, N. et al., "Adversarial classification", *Proc. ACM KDD '09*, Seattle, WA, August 2004.

[8] Roy, S. et al., "A survey of game theory as applied to network security", *Proc. HICSS 2010,* Honolulu, HI, January 2010.

[9] Williams, E., *Surveillance and Interdiction Models: A Game Theoretic Approach to Defend Against VBIED*, Thesis, Naval Postgraduate School, June 2010.

[10] Parameswaran, M., H. Rui, and S. Sayin, "A game theoretic model and empirical analysis of Spammer strategies", *Proc. CEAS 2010*, Redmond, WA, July 2010.

[11] Pita, J. et al., "GUARDS: Game theoretic security allocation on a national scale", *Proc. AAMAS '11*, Taipei, Taiwan, May 2011.

[12] Manshaei, M. et al., "Game theory meets network security and privacy", *ACM Computing Surveys,* December 2011.

[13] Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Second Edition, Springer, New York, 2009.

[14] *Trustworthy Cyberspace: Strategic Plan for the Federal Cybersecurity Research and Development Program,* December 2011.

[15] Verma, R. and D. Del Vecchio, "Safety control of hidden mode hybrid systems", *IEEE Trans. Automatic Control,* Vol. 57, pp. 62-77, 2012.

[16] Colbaugh, R., "Arctic ice, George Clooney, lipstick on a pig, and insomniac fruit flies: Combining kd and m&s for predictive analysis", *Proc. ACM KDD '11*, San Diego, CA, August 2011.

[17] Colbaugh, R. and K. Glass, "Predictive analysis for social processes I: Multi-scale hybrid system modeling, and II: Predictability and warning analysis", *Proc. 2009 IEEE MSC*, Saint Petersburg, Russia, July 2009.

[18] Sato, Y., E. Akiyama, and J.D. Farmer, "Chaos in learning a simple two-person game", *Proc. National Academy of Sciences USA,* Vol. 99, pp. 4748-4751, 2002.