A Network Security Classification Game[*]

Ning Bao[1], O. Patrick Kreidl[2], and John Musacchio[1]

[1] University of California Santa Cruz, Santa Cruz, CA 95064, USA
{nbao,johnm}@soe.ucsc.edu
[2] BAE Systems–Technology Solutions, Burlington, MA 01803, USA
pat.kreidl@baesystems.com

Abstract. We consider a network security classification game in which a strategic defender decides whether an attacker is a strategic spy or a naive spammer based on an observed sequence of attacks on file- or mail-servers. The spammer's goal is attacking the mail-server, while the spy's goal is attacking the file-server as much as possible before detection. The defender observes for a length of time that trades-off the potential damage inflicted during the observation period with the ability to reliably classify the attacker. Through empirical analyses, we find that when the defender commits to a fixed observation window, often the spy's best response is either full-exploitation mode or full-confusion mode. This discontinuity prevents the existence of a pure Nash equilibrium in many cases. However, when the defender can condition the observation time based on the observed sequence, a Nash equilibrium often exists.

Keywords: network security, classification game, sequential detection.

1 Introduction

In many network security situations, an operator of a network (the defender) may need to discern between different types of attackers. An attempt at espionage needs to be treated differently than an attack by a spammer for instance. Because of this, defenders will want to employ intrusion detection systems and related software to look for attack signatures and/or apply statistical tests. Knowing that a defender is trying to classify attacks, an attacker is likely to change the way they attack in order to make it more difficult to be classified correctly. These games can be quite complicated because they have both asymmetric information and they happen over time. In this work, we consider a very simple model of such a classification game, and with it we extract some key insights.

1.1 Basic Model

The model is illustrated in Fig. 1. The defender faces an attacker of two possible types: a spy or spammer with probabilities p and $1-p$ respectively. The defender has two servers that can be attacked, a File-Server (FS) and a Mail-Server (MS).

[*] Research supported by AFOSR grant FA9550-09-1-0049.

R. Jain and R. Kannan (Eds.): GameNets 2011, LNICST 75, pp. 265–280, 2012.

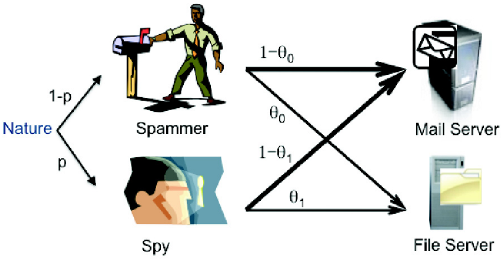


Fig. 1. An illustration of the classification game

We suppose that spammers attack the MS most often because they want to send spam and to get the addresses of potential victims. However, a spammer occasionally hits the FS as he explores the defender's network looking for other potential targets. We suppose time is discrete, and in each period k, a spammer hits the FS with probability θ_0 and otherwise he hits the MS. The attacks are restricted to be i.i.d. Bernoulli in each period. Moreover, we suppose the defender observes the sequence of attacks $z_k \in \{\mathrm{MS}, \mathrm{FS}\}$. Spammers are assumed to be non-strategic, so θ_0 is taken as a fixed parameter.

A spy chooses the frequency with which to hit the FS, which is the target with the information he wants. However, he can strategically choose to hit the MS during some time periods to make it more difficult for the defender to distinguish him from a spammer. We assume that the spy's single choice variable is θ_1, the probability of hitting the FS in any period. We restrict θ_1 to be larger than θ_0. By picking θ_1, the spy commits to attacking the FS according to a Bernoulli process with parameter θ_1. The spy's tradeoff is that if he picks θ_1 too high, then it is easy for the defender to distinguish him from a spammer, while if he lowers θ_1, he reduces the frequency with which he attacks his desired target.

The defender decides in each period whether to classify the attacker as a spammer or a spy, or to continue observing. While observing, the defender incurs a cost c_0 and c_1 for each MS hit by a spammer and FS hit by a spy, respectively, the latter a reward c_1 to the spy. The defender incurs a cost F if he mis-classifies a spammer as a spy. If he mis-classifies a spy as a spammer, we suppose that the spy can then continue to attack with impunity and thus earns a reward equal to the discounted net present value of an endless stream of FS attacks that happen with probability θ_1 in each period. This mis-classification reward to the spy, like the spy's rewards for all preceding FS attacks, appears as a cost to the defender.

1.2 Summary of Analysis and Results

We consider two versions of our classification game, differing in the class of strategies available to the defender. In the first version, the "commit to N" game (Sect. 3), the defender must commit to positive integer N, the number of observations he will take before making a classification. We show that at the end of N periods the defender should employ the well-known Likelihood Ratio Test

(LRT) [1], in which an optimal classification reduces to comparing the number of observed FS attacks to a certain threshold. The second version we call the "dynamic N" game (Sect. 4), as the defender can now decide in each period whether to continue to observe depending on what has been observed so far. Equivalently, the defender chooses a policy that maps observation sequences to control decisions (`continue`, `classify-spy`, `classify-spammer`). We show that the defender's best response to a spy's choice of θ_1 takes the form of the well-known Sequential Probability Ratio Test (SPRT) [2].

In both versions of our game, we focus on finding pure strategy Nash equilibria [3], all players playing pure strategies that are best responses to one another. Existence of such an equilibrium implies that it might be possible for the game to settle to a stable situation in which players behave predictably. In the specific context of our game, if we fix a defender strategy optimized for a particular hypothesis on θ_1 (i.e., if in the commit to N game we fix a particular observation window & LRT threshold or in the dynamic N game we fix a particular pair of SPRT thresholds), then it is possible that the spy's best response is to play with a θ_1 that does *not* match what the defender is expecting: a Nash equilibrium of the game would be a point where the spy's θ_1 and the defender's hypothesis $\hat{\theta}_1$ do match. Through a set of numerical experiments, in each case computing firstly the defender's best response to a hypothesis $\hat{\theta}_1$ and secondly the spy's best response θ_1 to that defender's strategy, we find that the commit to N game often has no pure Nash equilibrium whereas the dynamic N game often does.

1.3 Related Work

There is a growing body of work on attacker-defender security games, and much of it surveyed in [4]. Lye and Wing [5] propose a stochastic game model to study the behaviors of an administrator and an attacker in a Local Area Network (LAN). In a series of papers [6,7,8], Alpcan and Başar introduce a game-theoretic framework to model the interaction between the intrusion detection system (IDS) and attackers. Our game-theoretic framework focuses on attacker classification rather than intrusion detection. Our game also connects to the conclusions in [9], which showed that immediate expulsion is not the best response for all types of attackers. Statistical tests have been widely utilized in intrusion detection problems e.g., Jung and others [10] offer an on-line detection algorithm that identifies malicious port-scans using the seminal sequential hypothesis testing approach by Wald [2]. In the second version of our classification game, a similar but more general version of the Wald problem is studied and Wald's solution serves as the defender's best response function. Nelson and others [11] also study challenges to statistical classification when the defender faces a strategic attacker, but their focus is on vulnerabilities during the training of a classifier.

2 Detailed Model

This section formally describes the model introduced in Sect. 1, which involves prior probability $p \in (0, 0.5]$, the spammer's per-period probability $\theta_0 \in (0, 0.5]$

of a hit on the FS, and the spy's choice variable $\theta_1 \in (\theta_0, 1]$. Fig. 1 illustrates the situation. Also recall the positive-valued costs incurred by the defender: c_0 for each MS attack by a spammer, c_1 for each FS attack by a spy and F for mis-classifying a spammer as a spy. The positive-valued cost of mis-classifying a spy as a spammer will be expressed below in terms of other parameters.

2.1 Cost Functions of Defender and Spy

The cost functions of both players take the form of an expected total discounted cost with discount factor $\delta \in (0, 1)$. In the commit to N game, integer N is determined before play begins, and thus is not a function of the defender's observations. Because we assume the consequence of mis-classifying a spy is that the spy continues to attack the FS with a Bernoulli process of parameter θ_1, this mis-classification cost is $\sum_{k=N}^{\infty} \delta^k c_1\theta_1 = \delta^N c_1\theta_1/(1-\delta)$. Now define the following two (conditional) probabilities of making an error:

$$\alpha = \mathbf{P}\left[U = 1 \mid X = 0\right] \quad \text{and} \quad \beta = \mathbf{P}\left[U = 0 \mid X = 1\right], \tag{1}$$

where U denotes the classification decision of the defender after N observations and $X \in \{0, 1\}$ denotes the true type of the attacker being spammer or spy, respectively. It is standard [1] to call α the *false-alarm* rate (Type-I error probability) and β the *mis-detection* rate (Type-II error probability). Altogether, the defender's expected total discounted cost is given by

$$J^D = p\left\{\beta\delta^N \frac{c_1\theta_1}{1-\delta} + \sum_{k=0}^{N-1} \delta^k c_1\theta_1\right\} + (1-p)\left\{\alpha\delta^N F + \sum_{k=0}^{N-1} \delta^k c_0(1-\theta_0)\right\}, \tag{2}$$

which includes costs incurred due to (i) mis-detection of a spy, (ii) FS attacks by a spy, (iii) false-alarm of a spammer and (iv) MS attacks by a spammer. Similarly, the spy's expected total discounted cost is given by

$$J^A = -\beta\delta^N \frac{c_1\theta_1}{1-\delta} - \sum_{k=0}^{N-1} \delta^k c_1\theta_1, \tag{3}$$

which includes the reward of FS attacks (i) after the defender's mis-detection and (ii) before the defender's classification action.

When the defender's strategy is generalized to allow the observation sequence $(z_0, z_1, \ldots, z_k) \in \{\mathrm{MS}, \mathrm{FS}\}^{k+1}$ to influence when (and not just how) the classification is made, integer N becomes a random variable with distribution depending on both players' strategies. The two players' objectives are essentially the same as expressed by (2) and (3), but with the underlying expectations suitably generalized. Specifically, consider any positive integer n such that $\mathbf{P}\left[N = n\right]$ is nonzero and condition on the event that $(X, N) = (x, n)$. We first generalize the error probabilities in (1) to be $\bar{\alpha}(n) = \mathbf{P}\left[U = 1 \mid X = 0, N = n\right]$ and $\bar{\beta}(n) = \mathbf{P}\left[U = 0 \mid X = 1, N = n\right]$, respectively. We next generalize, for each

$k = 0, 1, \ldots, n-1$, the period-k probability of a FS attack by $\bar{\theta}_x(k|n) = \mathbf{P}\left[Z_k = \mathrm{FS} \mid X = x, N = n\right]$, which we note is not a constant by virtue of conditioning on the event $N = n$. We define

$$G_1^D(n) = \bar{\beta}(n)\delta^n \frac{c_1\theta_1}{1-\delta} + \sum_{k=0}^{n-1} \delta^k c_1 \bar{\theta}_1(k|n),$$

$$G_0^D(n) = \bar{\alpha}(n)\delta^n F + \sum_{k=0}^{n-1} \delta^k c_0 \left[1 - \bar{\theta}_0(k|n)\right],$$

$$G^A(n) = -\bar{\beta}(n)\delta^n \frac{c_1\theta_1}{1-\delta} - \sum_{k=0}^{n-1} \delta^k c_1 \bar{\theta}_1(k|n)$$

to be the defender's cost conditioned on the event $(X, N) = (1, n)$, the defender's cost conditioned on the event $(X, N) = (0, n)$ and the spy's cost conditioned on $N = n$, respectively. Altogether, the two players' cost functions are then defined by a final expectation over the stopping time N i.e.,

$$J^D = p \sum_{n=1}^{\infty} G_1^D(n)\mathbf{P}\left[N = n \mid X = 1\right] + (1-p) \sum_{n=1}^{\infty} G_0^D(n)\mathbf{P}\left[N = n \mid X = 0\right] \quad (4)$$

and

$$J^A = \sum_{n=1}^{\infty} G^A(n)\mathbf{P}\left[N = n \mid X = 1\right]. \quad (5)$$

2.2 Assumed Behavior of Spammer and Spy

Our model makes the assumption that either type of attacker is restricted to hitting his desired target according to a Bernoulli process. This is reasonable for a spammer, who is taken to be non-strategic and happens upon a FS rather than a MS simply by mistake. The restriction of the spy's strategy space to picking the rate of a Bernoulli process is indeed a simplifying assumption, but it has some justifications (and implications) that we now discuss.

First consider an alternative formulation in which the spy's strategy space were the set of all binary sequences (with 1 and 0 corresponding to FS and MS hits respectively). In equilibrium, if a spy were playing a mixed strategy, every sequence he assigns positive probability would have to have the same expected payoff. In our formulation with the strategy restricted to be Bernoulli, any finite sequence will have positive probability, but all sequences will not lead to the same payoffs. Therefore the restriction of the spy's choice to a single variable θ_1 – the parameter of the Bernoulli process – is structurally different than allowing the attacker to choose his sequence directly.

One interpretation is that the spy has a commitment device, like a computer program, that will pick the actual attack sequence once the spy has chosen a θ_1. The restriction of the attacker's strategy space to a single dimension greatly simplifies the game. Moreover the spy can make his attack much less predictable, and thus possibly more effective on average, with a commitment device like this. Another issue is how this commitment can be credible. One possible argument is that if the spy overrode his device when it picked an unfavorable sequence, he might increase his payoff on single attack but then become more predictable in future attacks. Making this argument formal is outside the scope of this paper.

As for using a Bernoulli process, suppose instead the commitment device were made to produce a different random process. In particular, suppose the spy considers any distribution for the commitment device so long as it achieves an expected number of FS hits of $N\theta_1$ over some period N. In any equilibrium (if it exists), the defender would eventually come to know the choice of distribution and play a best response to it. The defender could use this distribution in a likelihood ratio test of the spy vs. spammer hypotheses. The expected log-likelihood ratio, under the hypothesis that the attacker is a spy, is just the K-L divergence between the spy and spammer distributions. It turns out that this quantity is minimized, subject to his FS hit rate constraint, by choosing a Bernoulli process. To make this precise, let $B_N(\theta)$ denote the binomial distribution on length-N binary-valued sequences with success probability $\theta \in (0, 1)$.

Proposition 1. *Distribution $B_N(\theta_1)$ minimizes K-L divergence $D\left(P||B_N(\theta_0)\right)$ over P subject to the constraint that the expected number of successes is $N\theta_1$.*

The proof is in Appendix A.1. Proposition 1 addresses why a spy might want to use a Bernoulli process during the classification period, but it does not address the issue of why the spy would not change the process after the defender has made a classification decision. This is a more difficult question. For instance, if a spy could know he has just been misclassified, then he would want to change θ_1 to 1 and exploit the defender's mistake. But in game play, the spy would not know this had happened. If he were selecting a best response to a particular choice of N in the commit to N game, he could set his commitment device to increase θ_1 to 1 after time N. While we only allow a defender to classify once in our model, in a real situation a defender would likely reconsider the classification upon detecting such an abrupt change. A much more complex model would be needed to analyze this game. Therefore, we have elected to first understand a simpler model, limiting the spy's best response to a set of stationary policies.

3 Commit to N Game

In the commit to N game, the defender makes his classification decision after a fixed number N of observations, while the spy picks the probability θ_1 of hitting the FS in each attack. The problem (from the defender's point-of-view) turns out to be a standard binary hypothesis test for which we show that, for any N, the applicable Likelihood-Ratio Test (LRT) reduces to a comparison between the number of observed FS attacks and a certain threshold m. In turn, the defender's best response to a hypothesized strategy $\hat{\theta}_1$ is a pair of integers (N, m). Recall that under simultaneous-play assumptions the spy has no obligation (and, in fact, generally has incentives not) to behave as hypothesized by the defender.

A pure Nash equilibrium is a point for which the defender hypothesizes a value for $\hat{\theta}_1$, and designs his observation window N and LRT threshold m accordingly, such that the spy's best response to that integer pair (N, m) yields $\theta_1 = \hat{\theta}_1$.

3.1 Defender's Best Response

For any choice of N, the defender must decide between spammer or spy based on the observed sequence $z^N = (z_0, z_1, \ldots, z_{N-1})$ of server attacks. Given the spy's choice of probability θ_1 (and also values for model parameters p, θ_0, c_0, c_1, F and δ in (2)), this decision becomes equivalent to a binary hypothesis test between two binomial distributions $B_N(\theta)$: a spammer $H_0 : \theta = \theta_0$ versus a spy $H_1 : \theta = \theta_1$. The likelihood ratio is given by

$$L\left(z^N\right) = \frac{\mathbf{P}\left[z^N \mid X = 1\right]}{\mathbf{P}\left[z^N \mid X = 0\right]} = \left(\frac{\theta_1}{\theta_0}\right)^{\bar{z}} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{N-\bar{z}} \tag{6}$$

where $\bar{z}$ denotes the number of FS attacks in the given sequence z^N. By the Neyman-Pearson lemma, a decision rule of the form "reject H_0 if $L(z^N) > M$" subject to a Type-I error probability α is a level-α Uniformly-Most-Powerful (UMP) test [1], achieving for some M the minimum Type-II error probability β (and, in turn, a minimum in (2)) associated with the level α. It is easy to verify that, in the case of (6) with $\theta_1 > \theta_0$, the condition that $L(z^N) > M$ for any M is equivalent to the test $\bar{z} \geq m$ for some integer m. The following proposition provides, for any given N, the minimizing integer threshold m^* in closed form.

Proposition 2. *Fix strategy $\theta_1 \in (\theta_0, 1]$ for the spy. For any observation window N, the defender's optimal decision (with respect to minimizing (2)) is to classify the attacker as a spy if the observed sequence z^N contains a number of FS attacks*

$$\bar{z} \geq m^*(N) = \left\lceil \frac{\log\left(\frac{(1-p)F(1-\delta)}{pc_1\theta_1}\right) - N\log\left(\frac{1-\theta_1}{1-\theta_0}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right) - \log\left(\frac{1-\theta_1}{1-\theta_0}\right)} \right\rceil ;$$

otherwise, he classifies the attacker as a spammer.

The proof is in Appendix A.2. Recall that the defender's best response involves not just the choice of threshold m but also the choice of observation window N. Proposition 2 only tells us how to do the former given the latter, so the best response to a particular spy strategy θ_1 still involves a direct search on N. Moreover, in actual game play, the defender will not know the true value of θ_1. One could view this (for fixed N) as the defender carrying out a one-sided test $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. However, by properly choosing an alternative hypothesis $H_1' : \theta = \hat{\theta}_1$, the defender may effectively transform the one-sided test into a simple-vs-simple one. By the Karlin-Rubin theorem [1], the decision rule of Proposition 2 (for any given N) still achieves the smallest mis-detection rate β among all tests with the desired false-alarm rate α. Thus, for the purposes of looking for Nash equilibria, we can view the defender's best response strategy as being characterized by the hypothesis $\hat{\theta}_1$ for which the integer pair $(N(\hat{\theta}_1), m(\hat{\theta}_1))$ is an optimal choice.

3.2 Spy's Best Response

Consider the spy's best response $\theta_1 \in (\theta_0, 1]$ to a given defender's strategy, or integer pair (N, m). Substitution of the achieved mis-detection probability $\beta(N, m) = \sum_{i=0}^{m-1} \binom{N}{i} \theta_1^i (1-\theta_1)^{N-i}$ into (3) yields a spy's cost function that is an Nth-order polynomial in parameter θ_1. The minimizing argument will thus be either one of $N-1$ roots of the derivative polynomial or the boundary value of $\theta_1 = 1$, a total of up to N possibilities for J^A that can be compared numerically.

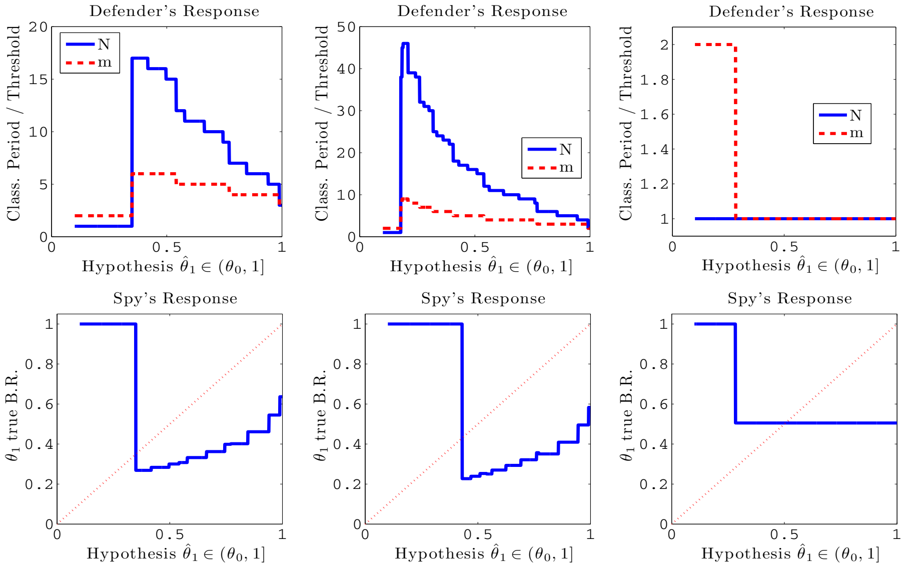
3.3 Numerical Experiments

The numerical procedure for each scenario is as follows. We search through a grid of possible $\hat{\theta}_1$ on the interval $(\theta_0, 1]$. For each $\hat{\theta}_1$, we search for the defender's best observation window N, always supposing that the LRT threshold m is chosen according to Proposition 2. This leads to a best response pair (N, m) against which we evaluate the spy's best response θ_1 via minimization of the Nth-order polynomial as described above.[1] Finally, we plot the best response θ_1 against $\hat{\theta}_1$, any point(s) where the two coincide being a Nash equilibrium.

Results for three representative scenarios are shown in Fig. 2, all assuming that (i) probability $p \leq 0.1$, or that spammers are much more common than spies, (ii) probability $\theta_0 = 0.1$, or that spammers mistakenly hit a FS only 10% of the time, (iii) costs $c_0 = 0.01$ and $c_1 = 1$, viewing a spied-FS much more costly than a spammed-MS, and (iv) discount factor $\delta = 0.99$ while cost $F \leq 10$, viewing a mis-detection at least as costly as a false-alarm. Each upper plot shows how the defender's strategy changes with his hypothesis $\hat{\theta}_1$, while each bottom plot shows the spy's best response θ_1 to that defender's strategy. In the first two scenarios, the spy's curve crosses the $\theta_1 = \hat{\theta}_1$ line at a discontinuity, and thus no Nash equilibrium exists; the third is a (not easily found) counter-example to any claim that a Nash equilibrium never exists in the commit to N game.

Examining Fig. 2 in more detail also exposes a confusion/exploitation tradeoff and interesting dynamics between defender and spy with respect to this tradeoff. Consider all three scenarios when $\hat{\theta}_1$ is near θ_0, or when the defender hypothesizes a spy favoring confusion over exploitation: the defender plays $(N, m) = (1, 2)$, corresponding to immediately choosing `classify-spammer`. That is, when the defender finds classification difficult and not worth the time expenditure required, he resorts to a trivial strategy that will ignore observations and rather base classification simply upon the relative costs between the two types of errors. The spy's best response, given that the defender is not even bothering to classify,

[1] If there are cases where there are multiple optimal pairs (N, m) for a given $\hat{\theta}_1$, it could be that some of these pairs support a Nash equilibrium and some do not. In our numerical experiments, however, we encountered no such cases.



(a) $p = 0.01$ and $F = 10$ (b) $p = 0.1$ and $F = 10$ (c) $p = 0.001$ and $F = 0.08$

Fig. 2. Numerical results for three scenarios of the commit to N game with parameters $\theta_0 = 0.1$, $c_0 = .01$, $c_1 = 1$ and $\delta = 0.99$, while parameters p and F vary as indicated. No Nash equilibrium exists in (a) or (b), but in (c) is a (rare) case with an equilibrium.

is to attack the FS at full rate. Next consider the scenarios as $\hat{\theta}_1$ increases up to where the defender first chooses to employ a non-trivial classification strategy. In the first scenario (with $p = 0.01$), the defender first chooses $N = 17$ at $\hat{\theta}_1 \approx 0.375$ and the spy's best response switches discontinuously to $\theta_1 \approx 0.269$ in order to allow some chance of being misclassified. In the second scenario (with $p = 0.1$) when the defender expects more spies, he first employs nontrivial classification at a lower $\hat{\theta}_1$ and, in turn, chooses a larger N to allow for adequate classification given that presumed stealthier spy; the spy chooses to exploit this long observation window and still attack at full rate, viewing the long stretch of FS attacks worth the sure detection–it is not until $\hat{\theta}_1 \approx 0.431$ that N is chosen small enough for the spy to make that first abrupt switch. In both scenarios, as $\hat{\theta}_1$ increases beyond the occurrence of the spy's first abrupt switch, the defender continues to choose a smaller N because under his hypothesis it is decreasingly difficult to discriminate between the two attackers; with this decreasing power of the defender's test, the spy is indeed choosing to increase his attack rate, but he continues to confuse by staying well below the increasing rates hypothesized by the defender. In the third scenario, when false alarms are very cheap, the first and only nontrivial classification strategy remains very simple as $\hat{\theta}_1$ increases: immediately classify spammer or spy in accordance with whether one attack is against the MS or FS, respectively. The spy, in turn, holds steady at $\theta_1 \approx 0.5$, giving rise to the equilibrium at $\theta_1^* \approx 0.5$ with $N(\theta_1^*) = m(\theta_1^*) = 1$.

A final point concerns the restriction that the spy picks θ_1 greater than θ_0. In the scenarios we discussed, this constraint was never active—the spy would not have picked $\theta_1 \leq \theta_0$ even if it were allowed. There are scenarios, particularly with larger θ_0, where the constraint could become active. Allowing $\theta_1 \in [0,1]$ only modestly complicates the game: the defender's best response to $\theta_1 < \theta_0$ still uses Proposition 2 but with the `classify-spy` decision made if $\bar{z} < m^*(N)$; in turn, the spy's best response to $\hat{\theta}_1 < \theta_0$ involves substitution of $\beta(N,m) = \sum_{i=m}^{N} \binom{N}{i} \theta_1^i (1-\theta_1)^{N-i}$ into (3), which remains an Nth-order polynomial in θ_1. Here (and in the next section), we considered only $\theta_1 > \theta_0$ for ease of exposition.

4 Dynamic *N* Game

In this section, we remove the restriction that the defender commits to a fixed observation window. As in the famous Wald problem [2], the number of observations N before classification depends not just on the two players' strategies but also on the particular observation sequence $z^N = (z_0, z_1, \ldots, z_{N-1})$. While our problem (from the defender's point-of-view) turns out to be a minor variation of the famous Wald problem, we show that the defender's best response, given θ_1 chosen by the spy, is in the form of a Sequential Probability Ratio Test (SPRT). The spy's choice variable is θ_1 just as it is for the commit to N game.

4.1 Defender's Best Response

Let us first show that, given the spy's choice of probability θ_1 (and also values for model parameters p, θ_0, c_0, c_1, F and δ in (4)), the defender can access a best response strategy in the family of Wald's SPRT solutions [2]. An SPRT strategy in each period k can be parametrized by two probability thresholds we will denote by $\eta_k \in [0,1]$ and $\xi_k \in [\eta_k, 1]$ i.e., initialize probability $b_{-1} = p$ and, in each period $k = 0, 1, 2\ldots$, first apply the probabilistic state recursion

$$\mathbf{P}\left[X = 1 \mid z^{k+1}\right] \equiv b_k = \begin{cases} \dfrac{(1-\theta_1)b_{k-1}}{(1-\theta_0)(1-b_{k-1}) + (1-\theta_1)b_{k-1}} & \text{, if } z_k = \mathrm{MS} \\[2ex] \dfrac{\theta_1 b_{k-1}}{\theta_0(1-b_{k-1}) + \theta_1 b_{k-1}} & \text{, if } z_k = \mathrm{FS} \end{cases} \tag{7}$$

and then choose to `classify-spammer` if $b_k \leq \eta_k$, to `classify-spy` if $b_k \geq \xi_k$, and to `continue` otherwise. Much is known for Wald's problem: for instance, under the criterion to minimize the expected infinite-horizon total cost, the optimal SPRT thresholds are stationary (i.e., neither η nor ξ varies its value with period k) [12].

The defender's problem in our model is similar to Wald's problem in most ways that general results in the field of stochastic dynamic programming have been organized i.e., both are infinite-horizon optimal stopping problems involving a partially-observable two-state system with bounded cost per stage [12]. One difference is that Wald's problem uses an expected total cost criterion without discounting. The other difference is that our single-stage cost of taking another observation also depends on the (unobservable) type of attacker. Even so, the defender's best response strategy remains in the set of stationary SPRTs.

Proposition 3. *Fix strategy $\theta_1 \in (\theta_0, 1]$ for the spy. A best response strategy for the defender (with respect to minimizing (4)) exists in the set of all stationary SPRT policies i.e., the set of all lower and upper thresholds $\eta \in [0,1]$ and $\xi \in [\eta, 1]$ applied in every period k to the probabilistic state b_k as described in (7).*

The proof is in Appendix A.3. While Proposition 3 gives us the form of the defender's response function, computing the actual SPRT thresholds via dynamic programming can only be done approximately because of the need to discretize the probabilistic state space $[0,1]$. A uniform discretization is always an alternative, but in certain problem instances a non-uniform discretization, favoring finer intervals in some sub-regions (e.g., around the concentration of the probabilistic state distribution, around the boundaries of the optimal thresholds), can significantly improve solution accuracy. Recognizing the defender's best response model as a special case of the well-studied Partially Observable Markov Decision Process (POMDP), we leverage a publicly available POMDP solver's implementation of such a non-uniform discretization.[2] Moreover, in actual game play, the defender will not know the true spy strategy but rather optimize SPRT thresholds and evolve the probabilistic state b_k via (7) based on a hypothesis $\hat{\theta}_1$.

4.2 Spy's Best Response

Consider the spy's best response to a given defender's strategy, or a given hypothesis $\hat{\theta}_1$ and the associated lower and upper thresholds $\eta(\hat{\theta}_1)$ and $\xi(\hat{\theta}_1)$ in the SPRT. A direct optimization of J^A in (5) over θ_1 is more complicated than was the case in the commit to N game. This is essentially because of the expectation with respect to N, whose distribution cannot be derived in closed form. Another complication is when the spy chooses $\theta_1 \neq \hat{\theta}_1$, the defender is not only employing sub-optimal SPRT thresholds but is also erroneously evolving his probabilistic state. The latter cannot be captured in the tractable dynamic programming reduction of Wald's problem.

Our approximation of the spy's best response is found by numerical searching over $\theta_1 \in (\theta_0, 1]$. For each value of θ_1, the spy's cost is found by constructing a finite-state Markov chain representation that exploits two key properties of the defender's SPRT strategy. Firstly, the probabilistic state recursion in (7) can (until classification) be equated with a random walk along the real line involving the defender's log-likelihood ratio (LLR)

$$R_k = \log\left(L(z^{k+1})\right) = \begin{cases} R_{k-1} + \log\left(\frac{1-\hat{\theta}_1}{1-\theta_0}\right) & \text{, if } z_k = \text{MS} \\ R_{k-1} + \log\left(\frac{\hat{\theta}_1}{\theta_0}\right) & \text{, if } z_k = \text{FS} \end{cases},$$

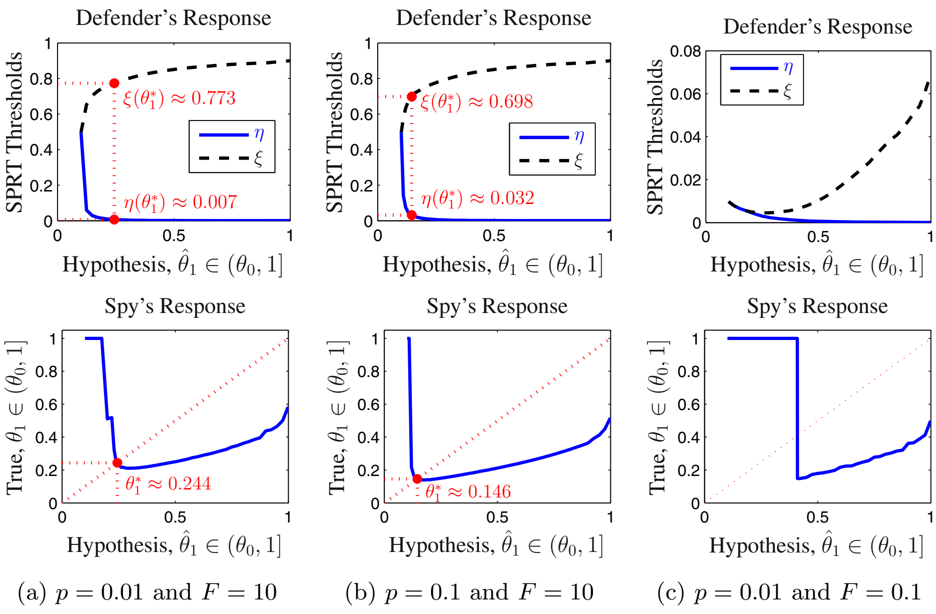
[2] Cassandra's implementation (see http://www.pomdp.org) of the "witness" algorithm [13] suited our setup particularly well.

starting from the origin $R_{-1} = 0$. In turn, the SPRT thresholds (and prior probability p) determine the segments of the real-line corresponding to the three control actions available to the defender i.e., choose to `classify-spammer` if $R_k \leq \log\left(\frac{(1-p)\eta(\hat{\theta}_1)}{p\left[1-\eta(\hat{\theta}_1)\right]}\right)$, to `classify-spy` if $R_k \leq \log\left(\frac{(1-p)\xi(\hat{\theta}_1)}{p\left[1-\xi(\hat{\theta}_1)\right]}\right)$, and to `continue` otherwise. Secondly, the spy's strategy θ_1 alters the statistics of this random walk (but *not* the increments $R_k - R_{k-1}$ themselves, which derive from $\hat{\theta}_1$), where lower (higher) values increase the chances that the LLR first exits the continue region at the lower (upper) end of the real-line. The Markov chain defines $Q + 3$ states, Q of them indexing the levels of a uniform quantization of the LLR continue region, one indexing an initial state and two indexing terminal states (one per classify decision). The transition probabilities reflect not only the true θ_1 and the increments $R_k - R_{k-1}$ of the defender's LLR walk based on hypothesis $\hat{\theta}_1$, but also the noise introduced by the quantization. The transition costs reflect the spy's rewards from file-server attacks and evading detection. A cost function is then defined on the state space by writing one-step calculations for each state. By solving this system of linear equations, we find the cost associated with the origin.

4.3 Numerical Experiments

The numerical procedure is analogous to that described for the commit to N game. For each hypothesis $\hat{\theta}_1$, we firstly employ the POMDP solver to obtain a particular SPRT threshold pair (η, ξ) for the defender and, secondly, employ the method described above to obtain a particular attack rate θ_1 for the spy. Any point(s) on the spy's response curve satisfying $\theta_1 = \hat{\theta}_1$ is (within the approximations discussed above) a Nash equilibrium. The results to be discussed used $Q = 100$ in the spy's response approximation and a precision of 0.001 to identify equilibria.

Fig. 3 shows results for three scenarios, the first two also considered for the commit to N game in Fig. 2. The key difference here is the smoother confusion versus exploitation tradeoff exhibited in the spy's best response curves. For example, let us compare the first scenario more closely. For hypotheses $\hat{\theta}_1$ closest to θ_0, the defender's response in both games are such that a spammer classification is made immediately regardless of the first observation; in turn, the spy's response is to hit the FS at every opportunity. However, in the dynamic N game, the defender's response first moves away from this trivial classification at a smaller $\hat{\theta}_1$ than in the commit to N game; at this point the spy's response also shifts away from his full rate but, different from the commit to N game, to rates $\theta_1 > \hat{\theta}_1$ that still exploit the confusion-oriented defense but not with full strength. For hypotheses $\hat{\theta}_1$ well away from θ_0, where an exploitative spy is anticipated, in both games the defender allows for time to reliably classify; in turn, the spy's response is to confuse (i.e., choose $\theta_1 < \hat{\theta}_1$) and better evade detection. Only the dynamic N game features a smooth transition between these two ends of play, at equilibrium $\theta_1^* \approx 0.244$ neutralizing the spy's incentive to either confuse an exploitation-oriented defense or to exploit a confusion-oriented defense.



(a) $p = 0.01$ and $F = 10$ (b) $p = 0.1$ and $F = 10$ (c) $p = 0.01$ and $F = 0.1$

Fig. 3. Numerical results for the dynamic N game in three scenarios, using the same parameters as in Fig. 2 (for the commit to N game) except for p and F in (c). A Nash equilibrium exists in (a) and (b), but in (c) is a (rare) case with no such equilibrium.

Comparison between the two games is similar in the second scenario (with larger p), finding in the dynamic N game a lower equilibrium point than that found in the first scenario. The third scenario is a (not easily found) counter-example to any claim that a Nash equilibrium always exists in the dynamic N game.

5 Conclusion

This work developed a classification game in the network security context. The defender tries to reliably classify the attackers (spammer or spy) while controlling the damage during evidence gathering. A strategic spy faces the trade-off between (i) exploiting the defender's observation time by attacking aggressively and (ii) confusing the defender by mixing attacks and enjoying the benefits of mis-detection. The frequent non-existence of pure Nash equilibrium in our commit to N game suggests that an over-simplified strategy adopted by the defender will prohibit the emergence of a stable situation where both players behave predictably. This problem is mitigated by allowing the defender to make decisions in every period as in our dynamic N game, which dis-incentivizes the spy's response from shifting drastically between exploitation mode and confusion mode.

Our game provides a new perspective to study the classification problem in network security, capturing many subtle yet important interplays between defender and attackers. The results to date are largely empirical, and we plan to explore a number of theoretical questions in future work. For the commit to N game, we desire a proof for the frequent non-existence of a pure Nash equilibrium, providing more insight to the limitations of this restriction on the defender's strategy; for the dynamic N game, it would be useful to isolate precise conditions on the game parameters for a pure Nash equilibrium to exist. Many model extensions are also possible, such as richer action spaces (e.g., sandboxing by the defender and disconnecting by the spy) or continuous-time variants.

Acknowledgment. The authors thank Gregory Frazier, Patrick Loiseau and Jean Walrand for conversations about this work, and Lemonia Dritsoula for help with the experiments.

References

1. Casella, G., Berger, R.: Statistical Inference. Duxbury Press (2002)
2. Wald, A.: Sequential Analysis. Wiley (1947)
3. Fudenberg, D., Tirole, J.: Game Theory. MIT Press (1991)
4. Alpcan, T., Başar, T.: Network Security: A Decision and Game-Theoretic Approach. Cambridge University Press (2011)
5. Lye, K., Wing, J.: Game Strategies in Network Security. Int. J. Information Security 4, 71–86 (2005)
6. Alpcan, T., Başar, T.: A Game Theoretic Approach to Decision and Analysis in Network Intrusion Detection. In: 42nd IEEE Conf. Decision and Control, pp. 2595–2600 (2003)
7. Alpcan, T., Başar, T.: A Game Theoretic Analysis of Intrusion Detection in Access Control Systems. In: 43nd IEEE Conf. Decision and Control, pp. 1568–1573 (2004)
8. Alpcan, T., Başar, T.: An Intrusion Detection Game with Limited Observations. In: 12th Int. Symp. Dynamic Games and Applications (2006)
9. Bao, N., Musacchio, J.: Optimizing the Decision to Expel Attackers from an Information System. In: Allerton Conf. on Comms., Control and Computing (2009)
10. Jung, J., et al.: Fast Portscan Detection Using Sequential Hypothesis Testing. In: IEEE Symp. Security and Privacy (2004)
11. Nelson, B., et al.: Misleading Learners: Co-opting Your Spam Filter. In: Tsai, J., Yu, P. (eds.) Machine Learning and Cyber Trust, pp. 17–51. Springer, Heidelberg (2009)
12. Bertsekas, D.P.: Dynamic Programming and Optimal Control, vols. 1 & 2. Athena Scientific (1995)
13. Kaebling, L., Littman, M., Cassandra, A.: Planning and Acting in Partially Observable Stochastic Domains. Artificial Intelligence 101, 99–134 (1998)
14. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)

Appendices

A.1 Proof of Proposition 1: The problem for sequences of length N is

$$\min_{P(\cdot)} \sum_{s\in\mathcal{S}} P(s)\log\left(\frac{P(s)}{\binom{N}{\bar{s}}\theta_0^{\bar{s}}(1-\theta_0)^{N-\bar{s}}}\right)$$

$$\text{s.t.} \sum_{s\in\mathcal{S}} P(s) = 1, \qquad \sum_{s\in\mathcal{S}} P(s)\bar{s} = N\theta_1, \qquad \text{and} \quad P(s) \geq 0 \;\; \forall s \in \mathcal{S}. \tag{8}$$

Here, $\bar{s}$ denotes the number of successes in any sequence s and S the set of all length-N sequences. This is a convex optimization problem (by convexity of K-L divergence) and the equality constraints are affine, in which case the KKT conditions are both necessary and sufficient for a global optimum [14]. Thus, distribution P is optimal if and only if

$$\log P(s) = \bar{s}\log(\theta_0) + (N-\bar{s})\log(1-\theta_0) + \mu + \lambda\bar{s} - \gamma_s - 1,\; \gamma_s \geq 0,\; \gamma_s P(s) = 0$$

and constraints (8) are met. The possible solution $P = B_N(\theta_1)$ with appropriately chosen KKT multipliers satisfies these conditions.

A.2 Proof of Proposition 2: With probability θ_1 and N fixed, each choice of integer m in a rule of the form $\bar{z} \geq m$ leads to a particular pair of error probabilities α_m and β_m. (For instance, $m = 0$ corresponds to "always classify as spy" and thus $\alpha = 1$ and $\beta = 0$.) To achieve values of α that do not correspond to integer m, one can introduce randomized decision rules, effectively mixing between the thresholds of m and $m+1$ for which $\alpha_{m+1} < \alpha < \alpha_m$. Thus, for a given N, the curve of achievable pairs of (α, β) (i.e., the "error curve") form a piecewise-linear curve. By writing the slope of each line segment explicitly, one can show that this error curve is also convex. Because the defender's objective J^D in (2) is linear in both α and β, the fundamental theorem of linear programming [14] implies that an optimal point always occurs at one of the vertices of the error curve. (In the degenerate case, when the slope of an error curve segment exactly matches the equal cost contours of J^D, a randomized decision rule can also be optimal, but gains nothing over the deterministic rule at either vertex).

The setting is illustrated in Fig. 4(a). Now consider the slope of a line perpendicular to the gradient of J^D (i.e., solve for $\frac{d\beta}{d\alpha}$ in the equation $\nabla_\alpha J^D = 0$). This slope can be no steeper (and no shallower) in magnitude than the slope of the error curve's segment to the left (and to the right, respectively) of m^*, so

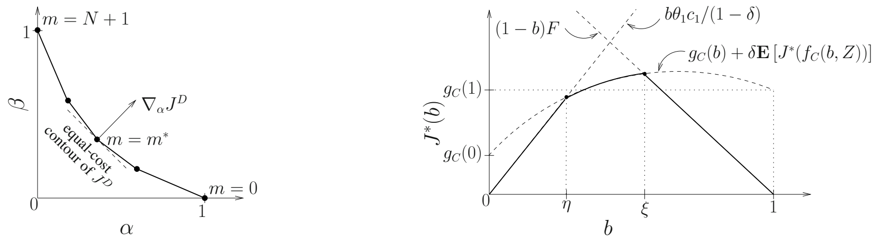
$$\left|\frac{\beta_{m^*} - \beta_{m^*-1}}{\alpha_{m^*} - \alpha_{m^*-1}}\right| \leq \frac{(1-p)F(1-\delta)}{p\theta_1 c_1} \leq \left|\frac{\beta_{m^*+1} - \beta_{m^*}}{\alpha_{m^*+1} - \alpha_{m^*}}\right|. \tag{9}$$

The log of the term on the right, after replacing m^* by m, is $m\log(\theta_1/\theta_0) + (N - m)\log((1-\theta_1)/(1-\theta_0))$. Equating this expression to the log of the middle term in (9), one can solve for an m that may not be integer. But with $\theta_1 > \theta_0$ this expression is monotone increasing in m, so taking $m^* = \lceil m \rceil$ insures that the right-side inequality of (9) is satisfied. Similar reasoning shows that this choice $m^* = \lceil m \rceil$ also satisfies the left-side inequality of (9).

A.3 Proof of Proposition 3: From Sect. 5.4 in Vol. 1 of [12], the imperfect state information problem involving a two-state system can be reduced to a perfect state information problem involving the probabilistic state recursion in (7). Like the partially-observable problem, the reformulated problem is an infinite-horizon discounted problem with bounded cost per stage, so from Sect. 1.2 in Vol. 2 of [12] the Bellman equation for all $b \in [0,1]$ specializes to

$$J^*(b) = \min\left\{ b\frac{\theta_1 c_1}{1-\delta}, (1-b)F, g_C(b) + \delta \mathbf{E}\left[J^*(f_C(b,Z))\right] \right\}$$

in which $g_C(b) = (1-b)(1-\theta_0)c_0 + b\theta_1 c_1$, the function f_C denotes the recursion of (7) and the expectation is with respect to the (mixed Bernoulli) distribution $\mathbf{P}[Z=z] = (1-b)\mathbf{P}[Z_k = z \mid X=0] + b\mathbf{P}[Z_k = z \mid X = 1]$. This is the same Bellman equation obtained for the standard infinite-horizon Wald problem (see Sect. 3.4 in Vol. 2) *except* that $g_C(b)$ is affine in b (rather than just a constant) and δ is not unity. It is easy to see that $J^*(0) = J^*(1) = 0$, and that $J^*(b)$ is bounded above by $\min\left\{b\frac{c_1\theta_1}{1-\delta}, (1-b)F\right\}$ on $[0,1]$. All arguments that lead to concavity of $J^*(b)$ on $[0,1]$ also still hold i.e., starting with $J^0(b) = 0$, function J^* is viewed as the point-wise limit of a sequence of functions $\{J^k\}$ resulting from repeated iterations of the Bellman equation in which (i) the monotonicity property of dynamic programming ensures that $J^k(b) \geq J^{k-1}(b)$ on $[0,1]$ and (ii) Sect. 5.5 in Vol. 1 ensures that if $J^{k-1}(b)$ is concave on $[0,1]$ then so is $\mathbf{E}\left[J^{k-1}(f_C(b,Z))\right]$ given f_C and $\mathbf{P}[Z=z]$ above, and thus so is $J^k(b)$. In turn, the reasoning to optimality of a stationary SPRT also still holds; see Fig. 4(b).



(a) (b)

Fig. 4. Illustrations of (a) the error curve discussed in Appendix A.2 and (b) the optimal cost-to-go function discussed in Appendix A.3