End-to-End Game-Focused Learning of Adversary Behavior in Security Games

1 Proof of Section 3 Theorems

Theorem 1 (Equal defender values)**.** *Consider a two-target SSG with a rational attacker, equal defender values for each target, and a single defense resource to allocate, which is not subject to scheduling constraints (i.e., any nonnegative marginal coverage that sums to one is feasible). Let z_0 and z_1 (w.l.o.g., we assume $z_0 \geq z_1$) be the attacker's values for the targets, which are observed by the attacker, but not the defender, and we assume w.l.o.g. are non-negative and sum to 1.*

The defender has an estimate of the attacker's values $(\hat{z}_0, \hat{z}_1)$ with mean squared error (MSE) ϵ^2*. Suppose the defender optimizes coverage against this estimate. If $\epsilon^2 \leq (1 - z_0)^2$, the ratio between the highest DEU under the estimate of $(\hat{z}_0, \hat{z}_1)$ with MSE ϵ^2 and the lowest DEU is:*

$$\frac{z_0 + \epsilon}{z_1 + \epsilon} \tag{1}$$

Proof. Given the condition that $\epsilon^2 \leq (1 - z_0)^2$, there are two configurations of $\hat{\boldsymbol{z}}$ that have mean squared error ϵ^2: $\hat{z}_0 = z_0 \pm \epsilon, \hat{z}_1 = z_1 \mp \epsilon$, yielding defender utility $-z_1 - \epsilon$ and $-z_0 - \epsilon$, respectively, because the attacker always attacks the target with underestimated value. The condition on $\epsilon^2 \leq (1 - z_0)^2$ is required to make both estimates feasible. $z_0 + \epsilon \geq z_1 + \epsilon$ because $z_0 \geq z_1$. □

Lemma 1. *Consider a two-target, zero-sum SSG with a rational attacker, and a single defense resource, which is not subject to scheduling constraints. The optimal defender coverage is $x_0 = z_0$ and $x_1 = z_1$, and the defender's payoff under this coverage is $-(1 - z_0)z_0 = -(1 - z_1)z_1$.*

Proof. The defender's maximum payoff is achieved when the expected value for attacking each target is equal, and we require that $x_0 + x_1 \leq 1$ for feasibility. With $x_0 = z_0$ and $x_1 = z_1$, the attacker's payoff is

$(1-z_0)z_0$ if he attacks target 0 and $(1-z_1)z_1 = (1-(1-z_0))(1-z_0) = z_0(1-z_0)$ if he attacks target 1. □

Theorem 2 (Zero-sum)**.** *Consider the same setting as Thm. 1 except the utilities are zero-sum. If* $\epsilon^2 \leq (1-z_0)^2$*, the ratio between the highest DEU under the estimate of* $(\hat{z}_0, \hat{z}_1)$ *with MSE* ϵ^2 *and the lowest DEU is:*

$$\frac{(1-(z_1-\epsilon))z_1}{(1-(z_0-\epsilon))z_0} \tag{2}$$

Proof. Given the condition that $\epsilon^2 \leq (1-z_0)^2$, there are two configurations of $\hat{\boldsymbol{z}}$ that have mean squared error ϵ^2: $\hat{z}_0 = z_0 \pm \epsilon, \hat{z}_1 = z_1 \mp \epsilon$, yielding defender utility $-(1-(z_1-\epsilon))z_1$ and $(1-(z_0-\epsilon))z_0$, respectively, because the attacker always attacks the target with underestimated value. The condition on ϵ^2 is required to make both estimates feasible. Because $z_0 \geq z_1$, $-(1-(z_0-\epsilon))z_0 \leq -(1-(z_1-\epsilon))z_1$. □

Theorem 3. *Consider the setting of Thm. 2, but in the case of a QR attacker. For any* $0 \leq \alpha \leq 1$*, if* $\lambda \geq \frac{2}{(1-\alpha)\epsilon} \log \frac{1}{(1-\alpha)\epsilon}$*, the defender's loss compared to the optimum may be as much as* $\alpha(1-\epsilon)\epsilon$ *under a target value estimate with MSE* ϵ^2*.*

Proof. Let $f(p)$ denote the defender's utility with coverage probability p against a perfectly rational attacker and $g(p)$ denote their utility against a QR attacker. Suppose that we have a bound

$$g(p) - f(p) \leq \delta$$

for some value δ. Let p^* be the optimal coverage probability under perfect rationality. Note that for an alternate probability $p' > p^*$

$$\begin{aligned} g(p') &\leq f(p') + \delta \\ &= f(p^*) - (p' - p^*)\epsilon + \delta \\ &\leq g(p^*) - (p' - p^*)\epsilon + \delta \quad \text{(since } f(p) \leq g(p) \text{ holds for all } p\text{)} \end{aligned}$$

and so any $p' > p^* + \frac{\delta}{\epsilon}$ is guaranteed to have $g(p') < g(p^*)$, implying that the defender must have $p' \leq p^* + \frac{\delta}{\epsilon}$ in the optimal QR solution.

We now turn to estimating how large λ must be in order to get a sufficiently small δ. Let q be the probability that the attacker chooses the first target under QR. Note that we have $f(p) = \epsilon p$ and $g(p) =$

$(1-p)(1-\epsilon)q + p\epsilon(1-q)$. We have

$$\begin{aligned} g(p) - f(p) &= (1-p)(1-\epsilon)q + p\epsilon(1-q) - \epsilon p \\ &= [(1-p)(1-\epsilon) - p\epsilon]q \\ &\leq q \end{aligned}$$

For two targets with value 1 and ϵ, q is given by

$$\frac{e^{\lambda(1-\epsilon)(1-p)}}{e^{\lambda\epsilon p} + e^{\lambda(1-\epsilon)(1-p)}} = \frac{1}{1 + e^{\lambda[\epsilon p - (1-\epsilon)(1-p)]}}$$

Provided that $\lambda \geq \frac{1}{\epsilon p-(1-\epsilon)(1-p)} \log\frac{1}{\delta} = \frac{1}{p-(1-\epsilon)} \log\frac{1}{\delta}$, we will have $g(p) - f(p) \leq \delta$. Suppose that we would like this bound to hold over all $p \geq 1 - \alpha\epsilon$ for some $0 < \alpha < 1$. Then, $p - (1-\epsilon) \geq (1-\alpha)\epsilon$ and so $\lambda \geq \frac{1}{(1-\alpha)\epsilon}\log\frac{1}{\delta}$ suffices. Now if we take $\delta \leq (1-\alpha)\epsilon^2$, we have that for $\lambda \geq \frac{2}{(1-\alpha)\epsilon}\log\frac{1}{(1-\alpha)\epsilon}$, the QR optimal strategy p' must satisfy $p' \leq 1 - \alpha\epsilon$, implying that the defender allocates at least $\alpha\epsilon$ coverage to the target with true value 0. Suppose the attacker chooses the target with value 1 with probability q^*. Then, the defender's loss compared to the optimum is $q^*\alpha\epsilon$. By a similar argument as above, it is easy to verify that under our stated conditions on λ, and assuming $\alpha \geq \frac{1}{2}$, we have $q^* \geq (1-\epsilon)$, for total defender loss $(1-\epsilon)\alpha\epsilon$. □

2 Section 4 Theorem

Theorem 4. *Let f be twice continuously differentiable and x be a strict local minimizer of f over $\mathcal{X}$. Then, at except on a measure zero set, there exists a convex set $\mathcal{I}$ around x such that $x^*_{\mathcal{I}}(\theta) = \arg\min_{x\in\mathcal{I}\cap\mathcal{X}} f(x,\theta)$ is differentiable. The gradients of $x^*(\theta)$ are given by the gradients of solutions to the local quadratic approximation $\min_{x\in\mathcal{X}} x^T\nabla^2 f(x,\theta)x + \nabla f(x,\theta)$.*

Proof. By continuity, there exists an open ball around x on which $\nabla^2 f(x,\theta)$ is negative definite; let $\mathcal{I}$ be this ball. Restricted to $\mathcal{X} \cap \mathcal{I}$, the optimization problem is convex, and satisfies Slater's condition by our assumption on $\mathcal{X}$ combined with Lemma 2. Therefore, the KKT conditions are a necessary and sufficient description of $x^*_{\mathcal{I}}(\theta)$. We now use this fact to give an explicit expression for the gradients of $x^*_{\mathcal{I}}(\theta)$. Since the equality constraints given by the function $\boldsymbol{h}$ are affine, we represent them as a matrix A, where $\boldsymbol{h(x)} = A\boldsymbol{x}$. The KKT conditions imply that $(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\nu})$ is an optimum if and only if the following equations hold:

$$\begin{aligned}
\boldsymbol{g}(\boldsymbol{x}) &\leq 0 \\
A\boldsymbol{x} &= 0 \\
\boldsymbol{\mu} &\geq 0 \\
\boldsymbol{\mu} \odot \boldsymbol{g}(\boldsymbol{x}) & \\
\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \theta) + \boldsymbol{\mu}^\top \nabla \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\nu}^\top A &= 0
\end{aligned}$$

Differentiating through this linear system using the implicit function theorem, as in Amos and Kolter (2017) and Donti et al. (2017), results in the following expression for the gradients of the optimal solution with respect to θ:

$$\begin{bmatrix} \frac{\partial \boldsymbol{x}}{\partial \theta} \\ \frac{\partial \boldsymbol{\mu}}{\partial \theta} \\ \frac{\partial \boldsymbol{\nu}}{\partial \theta} \end{bmatrix} = - \begin{bmatrix} \nabla_x^2 f(x, \theta) + \sum_{i=1}^{n_{ineq}} \mu_i \nabla_x^2 \boldsymbol{g}(\boldsymbol{x}) & \left(\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^T & A^T \\ diag(\boldsymbol{\mu}) \left(\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right) & diag(\boldsymbol{g}(\boldsymbol{x})) & 0 \\ A & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \nabla_x f(x, \theta)}{\partial \theta} \\ 0 \\ 0 \end{bmatrix} \tag{3}$$

□

We now note that the above expression depends only on the gradient and Hessian of f, along with the constraints g. Therefore, differentiating through the KKT conditions of the local quadratic approximation results in the same expression (since the quadratic approximation is defined exactly to be the second-order problem with the same gradient and Hessian as the original). This implies that $x_{\mathcal{I}}^*(\theta)$ is differentiable whenever the quadratic approximation is differentiable (i.e., whenever the RHS matrix above is invertible). Note that in the quadratic approximation, we can drop the requirement that $x \in \mathcal{I}$ since the minmizer over $x \in \mathcal{X}$ already lies in $\mathcal{I}$ by continuity. Using Theorem 1 of Amos and Kolter (2017), the quadratic approximation is differentiable except at a measure zero set, proving the theorem.

Lemma 2. *Let $g_1 ... g_m$ be convex functions and consider the set $\mathcal{X} = \{x : \boldsymbol{g}(x) \leq 0\}$. If there is a point x^* which satisfies $\boldsymbol{g}(x) < 0$, then for any point $x' \in \mathcal{X}$, the set $\mathcal{X} \cap B(x', \delta)$ contains a point x_{int} satisfying $g(x) < x_{int}$ and $d(x_{int}, x') < \delta$.*

Proof. By convexity, for any $t \in [0,1]$, the point $(1-t)x^* + tx'$ lies in $\mathcal{X}$, and for $t < 1$, satisfies $g((1-t)x^* + tx') < 0$. Moreoever, for t sufficiently large (but strictly less than 1), we must have $d((1-t)x^* + tx', x') < \delta$, proving the existence of x_{int}. □

3 Experiments

3.1 Experimental Setup

We run 60 trials per parameter combination. Unless it is varied in an experiment, the parameters are:

1. *Number of targets* $= |\mathcal{T}| \in \{8, 24\}$.
2. *Features per target* $= |\boldsymbol{y}|/|\mathcal{T}| = 100$.
3. *Number of training games* $= |D_{\text{train}}| = 50$. We fix the number of test games $= |D_{\text{test}}| = 50$.
4. *Number of attacks per training game* $= |\mathcal{A}| = 5$.
5. *Defender resources* is the number of defense resources available. We use 3 for 8 targets and 9 for 24.
6. We fix the attacker's weight on defender coverage to be $w = -4$, a value chosen because of its resemblance to observed attacker w in human subject experiments [1, 2]. All strategies receive access to this value, which would require the defender to vary her mixed strategies to learn.
7. *Historical coverage* $= \boldsymbol{p}_{\text{historical}}$ is the coverage generated by Unif, which is fixed for each training game.

3.2 Additional Graphs

The main purpose of these graphs is to show that 2S receives lower test entropy without regularization, but much worse defender expected utility. The regularization, while not beneficial from a prediction perspective, improves decision performance.

4 Significance and Standard Deviation

A paired sample t-test is used to measure significance.

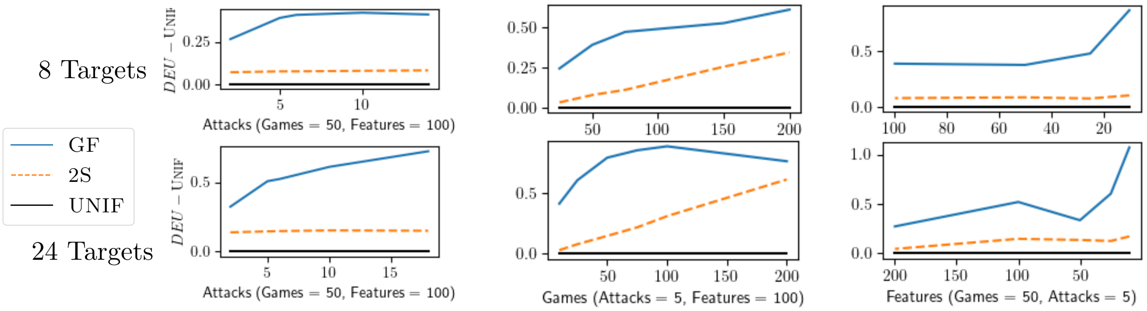

Figure 1: DEU − Unif for non-regularized 2S. 2S without regularization receives much worse decision performance than regularized 2S despite achieving lower test set cross entropy (see Fig. 2).

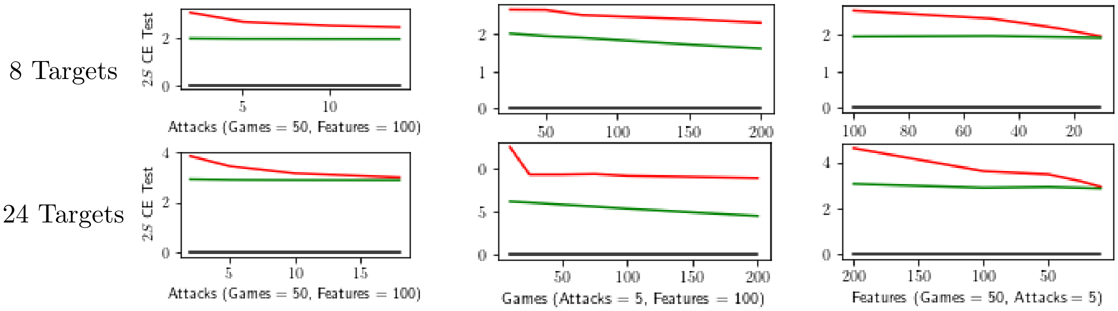

Figure 2: 2S test CE for non-regularized 2S (green) vs. regularized 2S (red). Regularization worsens test CE for 2S methods.

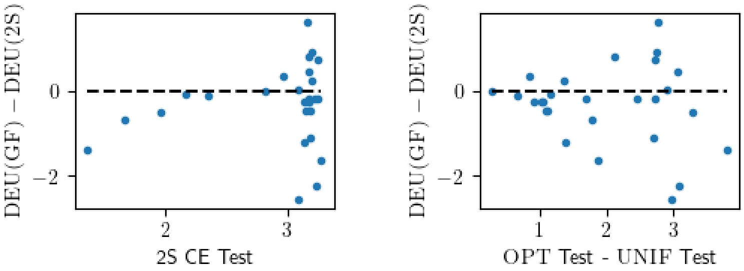

Figure 3: DEU(GF) − DEU(2S) with Dropout and early stopping based on a validation set for 24 targets, 100 features, 100 games and 5 attacks. Each point represents one trial. 2S receives worse test cross entropy compared to the non-regularized case, but better DEU.

4.1 Synthetic Data

Table 1: 8 Targets, Vary # of Attacks

Attacks	p-value	std 2S	std DF
2	3.8e-15	0.0109	0.0056
4	5.8e-13	0.0098	0.0064
6	8.7e-10	0.0088	0.0071
8	7.7e-06	0.0091	0.0073
10	0.018	0.0096	0.0074
12	0.22	0.0097	0.0068
14	0.87	0.0108	0.0077
16	0.57	0.0115	0.0079

Table 2: 8 Targets, Vary # of Games

Games	p-value	std 2S	std DF
25	0.033	0.0080	0.0075
50	4.0e-10	0.0100	0.0063
75	2.3e-06	0.0130	0.0080
150	0.0093	0.0160	0.0088

Table 3: 8 Targets, Vary # of Features

Features	p-value	std 2S	std DF
10	0.031	0.0153	0.0153
25	1.7e-4	0.0125	0.0107
50	0.65	0.0152	0.0099
100	4.0e-10	0.0100	0.0063
200	1.1e-29	0.0141	0.0058

Table 4: 24 Targets, Vary # of Attacks

Attacks	p-value	std 2S	std DF
2	0.0029	0.02116	0.0071
6	1.5e-4	0.0206	0.0099
10	0.030	0.0226	0.0094
14	0.89	0.0187	0.0120
18	0.75	0.0172	0.0127

Table 5: 24 Targets, Vary # of Games

Games	p-value	std 2S	std DF
10	0.32	0.0143	0.0087
25	0.62	0.0289	0.0124
50	1.1e-9	0.0291	0.0140
100	1.5e-4	0.0352	0.0148

Table 6: 24 Targets, Vary # of Features

Features	p-value	std 2S	std DF
10	0.42	0.0115	0.0110
25	2.9e-6	0.0120	0.0097
50	9.2e-10	0.0135	0.0088
100	1.6e-5	0.0207	0.0081
200	9.7e-48	0.0164	0.0051

4.2 Human-Subject Data

Table 7: 8 Targets, Vary # of Attacks

Attacks	p-value	std 2S	std DF
1	0.15	0.0070	0.0074
5	0.07	0.0063	0.0102
10	0.51	0.0074	0.0108
20	0.57	0.0070	0.0096
30	0.77	0.0070	0.0090

Table 8: 24 Targets, Vary # of Attacks

Attacks	p-value	std 2S	std DF
1	0.062	0.0067	0.0134
5	0.0012	0.0042	0.0127
10	0.0050	0.0071	0.0127
20	0.0040	0.0055	0.0117
30	0.0023	0.0052	0.0120

Table 9: 8 Targets, Vary # of Games

Games	p-value	std 2S	std DF
5	0.0021	0.0072	0.0039
10	0.88	0.0073	0.0061
15	0.24	0.0069	0.0081
20	0.038	0.0079	0.0095
25	0.24	0.0080	0.0106
30	0.38	0.0088	0.0093

Table 10: 24 Targets, Vary # of Games

Games	p-value	std 2S	std DF
5	2.2e-10	0.0049	0.0032
10	2.1e-4	0.0058	0.0062
15	4.7e-4	0.0069	0.0127
20	0.0059	0.0077	0.0123
25	0.025	0.0068	0.0102
30	0.074	0.0077	0.0124

References

[1] Thanh Hong Nguyen, Rong Yang, Amos Azaria, Sarit Kraus, and Milind Tambe. Analyzing the effectiveness of adversary modeling in security games. In *Proc. of AAAI-13*, Bellevue, Washington, 2013.

[2] Rong Yang, Benjamin Ford, Milind Tambe, and Andrew Lemieux. Adaptive resource allocation for wildlife protection against illegal poachers. pages 453–460, Paris, 2014.