A Model for Perimeter-Defense Problems with Heterogeneous Teams

Christopher D. Hsu\textsuperscript{1,2}, Mulugeta A. Haile\textsuperscript{1}, and Pratik Chaudhari\textsuperscript{2}

Abstract—We develop a model of the multi-agent perimeter-defense game to calculate how an adaptive defense should be organized. This model is inspired by the human immune system and captures settings such as heterogeneous teams, limited resource allocations, partial observability of the attacking side, and decentralization. An optimal defense, that minimizes the harm under constraints of the energy spent to maintain a large and diverse repertoire, must maintain coverage of the perimeter from a diverse attacker population. The model characterizes how a defense might take advantage of its ability to respond strongly to attackers of the same type but weakly to attackers of diverse types to minimize the number of diverse defenders and while reducing harm. We first study the model from a steady-state perimeter-defense perspective and then extend it to mobile defenders and evolving attacker distributions. The optimal defender distribution is supported on a discrete set and similarly a Kalman filter obtaining local information is able to track a discrete, sometimes unknown, attacker distribution. Simulation experiments are performed to study the efficacy of the model under different constraints.

I. INTRODUCTION

Consider two teams, denoted “defenders” (orange) and “attackers” (blue), each with a heterogeneous group of agents that have different capabilities, say offensive or defensive skills, or mobility. The defenders seek to protect an area against the attacking team which tries to penetrate this defense. The goal of this paper is to understand control policies for the defenders that take into account the heterogeneous nature of this problem.

We draw inspiration from the adaptive immune system which faces a similar problem while protecting an organism from pathogens\textsuperscript{1}. Some pathogens are persistent but benign (e.g., common cold) and others are rare but dangerous (e.g., HIV). There is a vast number of other pathogens on this spectrum. Tackling every pathogen optimally requires a specialized receptor (these are proteins expressed by lymphocytes whose molecules bind to the molecular constitution of pathogens). It would intuitively seem that the distribution of such receptors should be identical to the distribution of pathogens in the environment. But such a diverse repertoire would come at a large metabolic cost of maintaining all these receptors.

The immune system builds a rather non-intuitive repertoire: it devotes relatively more resources to fighting pathogens that cause large harm even if they are rare and relatively fewer resources to pathogens which cause less harm even if they are encountered frequently. Such a composition seems counter-intuitive. There are two key reasons which lead to its emergence. First, a constraint on the metabolic energy spent on maintaining a large receptor repertoire limits the total number of diverse receptors. And second, there exist specialized receptors that detect and tackle a specific pathogen with high probability but these receptors can also respond to other similar pathogens with a smaller probability and thereby reduce the overall harm to the organism. This paper embodies these ideas into control policies for the defender team.

A. Contributions

We use a theoretical model of the immune system\textsuperscript{1} to understand perimeter-defense problems with heterogeneous teams of multiple agents. The defender team seeks to select the appropriate type of agents and an appropriate number of them to minimize the harm caused by a heterogeneous attacker team. We seek to understand two situations: (i) when no single defender can defend against every type of attacker, and (ii) when defenders have limited resources that they should devote optimally to tackle the attackers.

We develop a technique to estimate an unknown attacker distribution. Defenders may not always have full knowledge of the attacker distribution. In this situation, it may be possible to use the information from defender-attacker encounters in a Bayesian filter to estimate the attacker distribution and adapt the defender distribution accordingly. We consider two situations: (i) when the attacker distribution is stationary and does not change over time, and (ii) when it evolves over time.

We show that the defender team can achieve near-optimal harm using a decentralized approach. Centralized computation may seem necessary to select the optimal defender distribution because both local and global structure of the attacker distribution determines it. We study decentralized dynamics where competition between defenders for successful interaction with attackers acts as a reward to encourage the congregation of specific defender types. We show how such local computation leads to near-optimal defender distributions, both when the attacker distribution is stationary, and when it evolves over time. We also study the “centralized estimation and decentralized control” setting where information obtained from individual interactions with the attackers is shared.

\textsuperscript{1}DEVCOM Army Research Laboratory christopher.d.hsu@army.mil, mulugeta.a.haile@army.mil

\textsuperscript{2}Department of Electrical & Systems Engineering and the GRASP Laboratory at the University of Pennsylvania. chsu8@seas.upenn.edu, pratikac@seas.upenn.edu
with all the defenders but the competitive dynamics of the defenders is decentralized.

B. Organization

We first formulate the model in §II. We investigate the model under different scenarios using theory and simulation experiments in §§III–VI. We discuss related work in §VII.

II. Problem Formulation

Let $Q_a$ be the probability that denotes the next intrusion will be caused by an attacker of type $a$. Let $P_d$ be the probability of a defender of type $d$ being present. We wish to compute the optimal distribution of the defenders $P_d^*$ that minimizes the harm caused by this intrusion. In biology, attackers $a$ and defenders $d$ interact in an abstract state-space called the shape-space $\mathbb{S}$.

In this paper, we will be deliberately vague about what the state-space is. For studying steady-state situations when there is a large number of attackers and defenders, we will think of the shape space as the types of agents (a defender of type $d$ interacts with an attacker of type $a$); for studying situations when defenders or attackers are mobile, we will think of the shape space as the Euclidean state-space $(a, d$ are locations of agents).

a) Interaction between attackers and defenders: Let us first model the situation when defenders are stationary on/inside the perimeter and get observations from the attackers when both are within a certain distance of each other. We will use a cross-reactivity term $f_{d,a}$ to denote the probability that defender $d$ tackles an attacker $a$ successfully. While the cross-reactivity can be any kernel, we will focus on a Gaussian $f_{d,a} \propto \exp \left( -\frac{(d-a)^2}{2\sigma^2} \right)$ where $\sigma$ denotes the bandwidth (as mentioned above, $\sigma$ can refer to the different types of agents, or Euclidean locations). Far away defenders under this kernel have a smaller chance of tackling attackers. The total probability of tackling an attacker $a$ successfully is $P_a = \sum_d f_{d,a} P_d$.

We can model the interaction between an attacker $a$ and some defender $d$ as a Poisson random variable with rate $\lambda_a(t)$. Larger the number of attackers, larger the rate of interaction with them; we model this as a rate that increases exponentially with time $\lambda_a(t) = \lambda_a \nu_a$ starting from some initial rate $\lambda_a(0) = 1$ (for sampled $a \sim Q_a$). In biology, this is because the population of pathogens grows. For us it is because the attacker spends more time in the environment.

The expected number of interactions of an attacker with some defender is $m_a(t) = \int_0^t d\tau \lambda_a(\tau) \approx \lambda_a(0)(\nu_a e^{\lambda_a(t)} - 1)/\nu_a$.

b) Harm caused by an attacker: The number of attackers of a type $a$ denoted by $m_a(t)$ also grows exponentially $m_a(t) = m_a \nu_a$ starting from some $m_a(0) = 1$ (for sampled $a \sim Q_a$) if it is not tackled successfully. The exponents $\nu_a$ and $\nu'_a$ may be different because the number of attackers $a$ that cause harm may grow differently than the number of attackers $a$ that can be observed by the defense. Unsuccessful interactions of defenders with attackers cause one unit of harm. For large times $t \approx \log(m_a/\lambda_a(0))/\nu_a$, the expected harm caused by an attacker of type $a$ after $m$ unsuccessful interactions is

$$F_a(m) = F_a(0) \left( \frac{m}{\lambda_a(0)} \right) \nu_a/\nu'_a \propto m^\alpha;$$

where $\alpha = \nu_a/\nu'_a$. The quantity $F_a(m)$ is polynomial in the number of interactions $m$ in this paper but it can also take other forms, e.g., $F_a(m) = 1 - e^{-\beta m}$ would model the situation where the harm plateaus after a large number of interactions. The harm $\bar{F}_a$ caused by an attacker $a$ is thus the average recognition time.

$$\bar{F}_a = \bar{F}_a = \int_0^\infty dm F_a(m) e^{-mP_a}.$$  \hspace{1cm} (2)

We call this the “empirical harm” because we will be able to estimate it using simulations.

c) Minimizing the harm: The harm caused by an attacker sampled from $Q_a$ is

$$\text{Harm}(P_d) = \sum_a Q_a \bar{F}_a.$$  \hspace{1cm} (3)

It is a function of the defender distribution $P_d$ and our goal will be to minimize it. When $F_a(m) = m^\alpha$, the optimization problem can be solved analytically to obtain the mean harm

$$\bar{F}_a = \frac{\Gamma(1 + \alpha)}{P_a^\alpha},$$  \hspace{1cm} (4)

where $\Gamma$ is the Gamma function; see the Appendix. We will refer to this as the “analytical harm” in the rest of the paper. The harm incurred for different settings, e.g., a sub-optimal defender distribution $P_d$, can be compared to this quantity.

III. Optimal Defender Distribution is Supported on a Discrete Set

Our cross-reactivity term models the fact that in biology, a given receptor can bind to different pathogens to varying degrees: a non-zero bandwidth $\sigma$ allows the organism to reduce the number of different types of defenders that it needs to maintain to tackle typical attackers. For a Gaussian $Q_a \propto e^{-\sigma^2/(2\sigma^2)}$, shows that when the cross-reactivity bandwidth $\sigma$ is smaller than a threshold $\sigma_Q \sqrt{1 + \alpha}$, the optimal defender distribution is also Gaussian

$$P_d^* \propto e^{-\frac{(1 + \alpha) \sigma^2}{2 \sigma_Q^2}}.$$  \hspace{1cm}

Note that the variance is negative if $\sigma/\sigma_Q > \sqrt{1 + \alpha}$. So if the bandwidth is above this threshold, the optimal defense is a Dirac delta at the origin. This shows that cross-reactivity allows concentrated distributions and reduces the need to maintain a diverse number of defenders. We can also show this mathematically where we can turn the variational problem in (3) into a standard optimization problem by assuming a parametric form for the defender distribution $P_d = e^{-\sigma^2/(2\sigma^2)}$ (by symmetry it has to be centered at $d = 0$). For a Gaussian $Q_a$, we can now minimize the harm in (3), to see that the solution has $\sigma_P = 0$. 


Such discreteness is also seen in many other problems, e.g., capacity of a Gaussian channel under an amplitude constraint\(^5\) and representations in the brain\(^4\). A somewhat related phenomenon can also be found in the statistics literature which studies priors that are supported over a discrete set\(^5\). In intuitive terms, a discrete prior (or a discrete defender distribution) allows capturing large parts of the data space (or the attacker distribution) with few resources (models or receptors). These ideas have been studied in information theory\(^6\) and have also been used to build new methods to learn from unlabeled data in the machine learning literature\(^7\).

In our case, the optimal defender distribution \(P_d^*\) also puts its probability mass on diverse types of defenders to best tackle the attackers using the cross-reactivity. Larger the cross-reactivity bandwidth \(\sigma\), more the spacing in between the atoms of \(P_d^*\). Although, we have argued discreteness in the setting when \(Q_a\) is Gaussian, this result is expected to hold in general; see the numerical simulation in Fig. 1.

### A. A toy perimeter-defense problem with immobile attackers and defenders

We first discuss how cross-reactivity helps tackle heterogeneity using a toy problem where neither defenders nor attackers can “move”. Imagine that a fixed number of attackers spawn according to the attacker distribution \(Q_a\) and move with constant velocity towards a perimeter. Similarly, a fixed number of defenders spawn on the perimeter according to the defender distribution \(P_d\). For a 1-dimensional perimeter defense problem, let us have the defenders be on the real line and they can only interact with attackers who seek to penetrate the line; see Fig. 1. The shape space here is the Euclidean location of the attackers and defenders. Defenders that are close to attackers are able to interact with the attacker with a higher probability and if the interaction succeeds, they can prevent harm. On the other hand, defenders that are very far away from the attacker have a small probability of successfully interacting with the attackers and are thus unable to stop the harm. The goal of the defense team is to place defenders along the perimeter and minimize the number of attackers that successfully penetrate the perimeter. This setting can be exactly captured by our model; the optimal defender distribution is shown in Fig. 1.

### IV. Modeling mobile defenders

We now formulate the problem where defenders can move to different parts of a Euclidean state-space to intercept attackers. Again the coordinates \(a, d \in \mathbb{R}\) are in Euclidean space. For the perimeter-defense setup, defenders begin at location \(d\) and can move on the real line with a speed \(u\). Attackers arrive at a location \(a\). Only the relative speed of the attackers and defenders matters, so we can set the speed of the attackers to be 1. Our key idea is to model the movement of the defenders using the cross-reactivity \(f_{d,a,u}\). Consider the scenario where the intruder enters the field of view of the sensors of the defenders when it is 1 unit away from the

![Fig. 1: Top: Defenders protect a perimeter [0, 1] from attackers with distribution \(Q_a\) supported on [0, 1]. Probabilities \(Q_a\) are drawn from a log-normal distribution with coefficient of variation \(\kappa^2 = \exp(\sigma_a^2) - 1\) and normalized (blue). The optimal defender distribution \(P_d^*\) (orange) is found by optimizing (3). A cross-reactivity kernel \(f_{d,a}\) of bandwidth \(\sigma = 0.05\) leads to a discrete distribution that covers the state space. In green, we show the harm \(\bar{F}_a\) caused by attackers at different locations \(a\) weighted by their probability \(Q_a\). This harm is relatively constant in spite of a discrete distribution \(P_d^*\) because of the cross-reactivity kernel \(f_{d,a}\) with a non-zero bandwidth \(\sigma\). The number of spikes in \(P_d^*\) will reduce for a larger \(\sigma\) and although the overall harm \(\sum Q_a \bar{F}_a\) will increase, the harm by every attacker \(\bar{F}_a\) will be relatively constant.

Bottom: The harm incurred using a non-optimal \(P_d\) is close to that of the harm of \(P_d^*\) in Fig. 1 even for large differences, as measured in the total variation norm \(\frac{1}{2} \|P_d - P_d^*\|_1\). To obtain this plot, we sampled 1000 different \(P_d\)s (by perturbing the optimal \(P_d^*\) using log-normal noise) and computed the empirical and analytical harm against a fixed \(Q_a\). This experiment also indicates that the empirical harm calculated using our experiments using (2) is close to the analytical harm (4) for a broad regime.
perimeter. We set
\[ f_{d,a,u} \propto e^{-(d-a)^2/(2u^2\sigma^2)} \]
If \( u \ll (d-a)/(2\sigma) \), then the probability of interception is small and harm occurs. Our goal is to design the optimal distribution for the defenders \( P_{d,u}^* \) that minimizes the harm caused by the attackers who cannot be intercepted. This setup can be solved using the same problem formulation as that of §II now by setting \( P_a = \sum_{d,u} f_{d,a,u} P_{d,u} \). Fig. 2 discusses simulation experiments with mobile defenders.

V. ESTIMATING AN UNKNOWN ATTACKER DISTRIBUTION

Next we are interested in understanding settings where the attacker distribution \( Q_a \) is not known to the defenders and it may change over time. The attacker distribution can change in two ways, either there are different types of attackers that arrive as time progresses, or the attackers move in the state-space. We will discuss how to estimate \( Q_a(t) \) and use this estimate to calculate a new optimal distribution for the defenders \( P_d(t) \) at each time step. Interactions, both successful recognition events or unsuccessful ones, give observations of the attacker types \( a \) and thereby \( Q_a(t) \).

Our model for how attackers proliferate enables us to estimate their distribution as follows. Let us first assume that \( Q_a \) is unknown to the defenders but it is stationary, i.e., it does not change with time. From §II, the expected number of attackers is \( m_a(t) \) and it grows as \( m_a(t) = m_a(0)e^{\alpha t} \). Since we have \( m_a(0) = 1 \) for attackers that were sampled from \( Q_a \), if we can estimate \( m_a(0) \) then we can think of the attackers corresponding to it as sampled from \( Q_a \) and therefore obtain an estimate of \( Q_a \).

Let us arrange all the attackers into a large vector \( x(t) \in \mathbb{R}^{m_a(t)} \) which will be the state of the filter. We will assume that \( x_i \sim N(\mu_i, \sigma_i) \) for \( i \leq m_a(t) \); observations \( y_i = cx_i + \xi \) where \( c \) is a Boolean that indicates if this specific attacker interacted with the defenders; \( \xi \sim N(0, \sigma_\xi^2) \) is Gaussian noise. We can update \( \mu_i^+ = \mu_i + k(y_i-c\mu_i) \) and \( (\sigma_i^+)^2 = (1-kc)^2\sigma_i^2 \), using the Kalman gain \( k = \sigma_i^2/c(\sigma_i^2 + \sigma_\xi^2) \). The growth rate is estimated as \( \hat{\nu}_a = \log(\mu_a^+ / \mu_a) \). Since there are multiple attackers for each type, we average the estimates as
\[
\hat{\mu}_a(t) = \frac{\sum_{i=1}^{m_a(t)} \mathbf{1}_{\{\text{type } i = a\}} \mu_i}{\sum_{i=1}^{m_a(t)} \mathbf{1}_{\{\text{type } i = a\}}} ,
\]
and similarly for \( \hat{\nu}_a \) to get \( \hat{Q}_a(t) = \hat{\mu}_ae^{-\hat{\nu}_a t} \). Let us emphasize that the number of attackers \( m_a(t) \) (and therefore the dimensionality of the state \( x \) of the filter) growing with time is not an issue in the implementation. As discussed, the filter runs independently for each attacker (and therefore each attacker type) but is easy to incorporate a known model of the correlations between the growth rates \( \nu_a \) of different attacker types. Given the estimate of \( \hat{Q}_a(t) \), we recalculate the optimal defender distribution \( P_d(t) \) by minimizing (3) at each instant. Fig. 3 discusses simulation experiments with a stationary \( Q_a \).

The above expressions for \( \hat{Q}_a(t) \) are derived for a stationary
When the last attacker is recognized by the defense and total
the corresponding defender distributions
incurs the optimal harm $N$
attacker distribution and obtain a defender distribution that
in subsequent episodes is at most twice the optimal $N$

Due to noisy observations resulting from suboptimal defender
distributions which further result in an inaccurate estimate of
$harm of the episode is the harm incurred until all attackers are
recognized, the estimate $Q_\alpha$ decreases with the number of episodes
$a (t)$, which is a decreasing function of its argument, the quantity
is weighted by some function $\varphi(\cdot)$ which is a decreasing function of its argument. The quantity
determines how many other defenders can tackle $a$. If this is large, then the incremental utility of having the
specific defender $d$ diminishes, and therefore its growth rate $\nu_d$ should be smaller. We will see in experiments that the
harm caused by this decentralized dynamics converges to the
optimal harm obtained from a centralized computation of $P^*_d$. We can also use the estimated attacker distribution $Q_a$ in place of the true distribution $Q_a$ in (6) situations where it is unknown to the defenders.

Fig. 5 shows an experiment with decentralized control when the
attacker distribution $Q_a$ is stationary while Fig. 6 shows a similar result when defenders estimate an evolving attacker
distribution using the filter in §V. We have also obtained similar results for the case when defenders move and tackle a
stationary attacker distribution (like that of §IV).

VII. Discussion and Related Work

Pursuit-evasion games have seen a wide range of perspectives, surveys, applications such as search and rescue, and environmental monitoring. Problems with multiple agents are complex due to their high dimensional
HpletL orangeI converges to the analytical harm HpletL greenI plotsI and the estimated harm HpletL blueI right plotsI for two situationsL the peaks of the estimated harm are close to optimal harm which is calculated using the true distribution Qa( = 1) immediately but models the growth of defenders over time. We run the population dynamics for 10^6 iterations within each episode to obtain Pd(t). The defender distribution (orange, right) is again close to the optimal distribution Pd for Qa but note that it is a bit smoother (and therefore sub-optimal). We know from Fig. 2 that there is a tradeoff between the diversity of the defender types (many types for a smooth distribution to few types for a spiked distribution) and harm (lower harm for the smooth distribution and higher harm for same recognition probability for the latter).

Fig. 5: Convergence to near-optimal harm using decentralized control. For the same setting as that of Fig. 3, we run the dynamics in (6) using the estimate Qa(t) calculated using the Kalman filter to induce the optimal defender distribution Pd∗(t) (instead of calculating it by minimizing the harm in (3) as done in Fig. 3). The harm in this case is slightly lower than that in Fig. 3, perhaps because of the transient dynamics in (6) which does not lead to Pd∗ immediately but models the growth of defenders over time. We run the population dynamics for 10^6 iterations within each episode to obtain Pd(t). The defender distribution (orange, right) is again close to the optimal distribution Pd for Qa but note that it is a bit smoother (and therefore sub-optimal). We know from Fig. 2 that there is a tradeoff between the diversity of the defender types (many types for a smooth distribution to few types for a spiked distribution) and harm (lower harm for the smooth distribution and higher harm for same recognition probability for the latter).

Fig. 4: Estimating a changing attacker distribution. For the same setting as that of Fig. 3, we run the filter when the attacker distribution shifts as Qa+κ = Qa, i.e., after the kth episode the probability of observing attacker a+i is Qa. The interval [0,1] is set to be isomorphic to [0,2π] in this case by setting a = 0 and a = 1 to be the same points. If the shape space consists of Euclidean locations of the agents, this models mobile attackers.

Top: Even if the estimation scheme is designed for a stationary Qa, it can track an evolving attacker distribution (left). The initial (κ = 0, blue) and final (κ = 300, orange) attacker distributions Qa for coefficient of variation κ = 1 and the final estimation Qa (green) at k = 300 (right). This shows that it is easier to estimate an evolving attacker distribution with a smaller κ.

Middle (κ = 1) and Bottom (κ = 3): We show the final Qa (blue, right plots) and the estimated Pd∗(Qa) (orange, right plots) for two situations, κ = 1 (middle) and κ = 3 (bottom). In both cases, the peaks of the the estimated Pd∗ track the peaks of Qa. The empirical harm incurred in this experiment (left, orange) converges to the analytical harm (left, green) and is close to optimal harm (left, blue) which is calculated using the true Qa at each time step. The superior tracking for κ = 1 leads to lower variance in the empirical harm (middle left, orange) and a more biased Pd∗ (middle right, orange) that is catered to the final Qa(κ = 1). In contrast, a poorer estimate of Qa(κ = 3) (top left, orange) leads to a larger variance in the empirical harm (bottom left, orange). Note how the attacker distribution (blue, middle-bottom right) has more spikes for a larger κ. This experiment shows how cross-reactivity allows the defenders to both track and counter different types of attacker distributions.

A theme that has emerged recently in reinforcement learning-based multi-agent control is to obtain decentralized, cooperative policies to tackle non-stationarity and limited information. A popular paradigm is to train a centralized policy and execute it in a decentralized fashion. Emergence of coordination has also been studied. These methods typically suffer from poor sample complexity but some have been scaled to large problems using heuristics.

This work uses non-learned based methods but demonstrates scalable decentralized policies that can adapt to a wide range of changing attacker distributions for problems as large as 1000 attackers and defenders, albeit in the limited context of a perimeter-defense game. We also obtain a useful guiding principle: competition among the defenders where successful interaction with the attackers acts as the reward may achieve optimality of the resources spent. Such ideas can be used for designing new reinforcement learning methods for multi-agent control. Prior work has noticed the benefits of a similar competition arising out of stochastic policies.

state-space. We investigate a variant of the pursuit-evasion game first introduced as the target-guarding problem where the pursuerndefender tries to prevent the evaderintruder from reaching the target. A large body of the work on this problem studies how multiple defenders decompose the problem into smaller games, or reduce the defense strategy to an assignment problem.

A theme that has emerged recently in reinforcement learning-based multi-agent control is to obtain decentralized, cooperative policies to tackle non-stationarity and limited information. A popular paradigm is to train a centralized policy and execute it in a decentralized fashion. Emergence of coordination has also been studied. These methods typically suffer from poor sample complexity but some have been scaled to large problems using heuristics.

This work uses non-learned based methods but demonstrates scalable decentralized policies that can adapt to a wide range of changing attacker distributions for problems as large as 1000 attackers and defenders, albeit in the limited context of a perimeter-defense game. We also obtain a useful guiding principle: competition among the defenders where successful interaction with the attackers acts as the reward may achieve optimality of the resources spent. Such ideas can be used for designing new reinforcement learning methods for multi-agent control. Prior work has noticed the benefits of a similar competition arising out of stochastic policies.
Our work is inspired from how the adaptive immune system in humans is organized. We strongly build upon the recent literature in biophysics which models the response of the adaptive immune system in an environment with a large number of evolving pathogens\cite{12,14,15}. We believe that the salient traits exhibited in the immune system such as heterogeneous defender types with wide cross-reactivity that interact with a range of attackers, and decentralized estimation and execution are key building blocks for understanding multi-agent autonomy. There are also a number of works that use bio-inspired approaches mimicking the behavior of ants, bees and flocks\cite{26,27,28} for multi-agent control.

Broadly speaking, modeling heterogeneity has received relatively little interest\cite{29}. Heterogeneity comes in many forms such as differences in roles\cite{30,31,32}, robotic capabilities and or sensors\cite{33,34}, dynamics\cite{35,36}, and even teams of air and ground robots\cite{37,38}. But algorithmic methods that can tackle large-scale heterogeneity have been difficult to build. This work uses a simple formulation to understand how to devote resources optimally in a heterogeneous environment.

**Fig. 6: Convergence to near-optimal harm using decentralized control for the case when the attacker distribution evolves across episodes.** For the same setting as that of Fig. 4, i.e., when the attacker distribution shifts as $Q_{a+k} = Q_{a}$, we implemented the decentralized control dynamics using the estimate $\hat{Q}_{a}$ obtained by the defenders. We have implemented a centralized estimation problem in this case, i.e., all defenders share information of their interactions with each other and thereby maintain a single estimate $Q_{a}$.

**Top:** We show $Q_{a}$ with coefficient of variation $\kappa = 3$ for the initial ($k = 0$) and final ($k = 300$) attacker distributions and the final estimation $\hat{Q}_{a}$ at $k = 300$ (right). Like Fig. 4, the TV plot (left) shows that estimation of $Q_{a}$ is easier for a smaller $\kappa$.

**Middle ($\kappa = 1$) and Bottom ($\kappa = 3$):** Like Fig. 5, the decentralized formulation gives rise to smooth defender distributions (orange, right). In the middle row, the distribution $Q_{a}$ is less spiked and therefore the decentralized defender distribution is more spiked (and closer to the optimal one, as indicated by the empirical harm in orange vs. analytical harm in blue in the middle left plot). In the bottom row, the attacker distribution $Q_{a}$ is more spiked which makes it more difficult to be estimated by the filter. Consequently, the defender population induced by the decentralized dynamics is also sub-optimal (orange in bottom right is farther away from optimal, as indicated by the empirical harm in orange being farther away from the optimal harm in blue in the bottom left plot). The empirical harm for $\kappa = 3$ again has a higher variance than that of the middle row. This experiment indicates that the decentralized population dynamics in (6) is capable of both tracking and countering an evolving attacker distribution.

**References**

\[1\] Mayer, A., Balasubramanian, V., Mora, T. & Walczak, A. M. How a well-adapted immune system is organized. *PNAS* **112**, 5950–5955 (2015).

\[2\] Perelson, A. S. & Oster, G. F. Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-non-self discrimination. *Journal of Theoretical Biology* **81**, 645–670 (1979).

\[3\] Smith, J. G. The information capacity of amplitude-and variance-constrained scalar Gaussian channels. *Information and control* **18**, 203–219 (1971).

\[4\] Laughlin, S. A simple coding procedure enhances a neuron’s information capacity. *Zeitschrift für Naturforschung c* **36**, 910–912 (1981).

\[5\] Berger, J. O., Bernardo, J. M. & Sun, D. The formal definition of reference priors. *The Annals of Statistics* **37**, 905–938 (2009).

\[6\] Mattingly, H. H., Transtrum, M. K., Abbott, M. C. & Machta, B. B. Maximizing the information learned from finite data sets: Simple models. *PNAS* **115**, 1760–1765 (2018).

\[7\] Gao, Y., Ramesh, R. & Chaudhari, P. Deep reference priors: What is the best way to pretrain a model? *ICML* (2022).

\[8\] De Boer, R. J. & Perelson, A. S. T cell repertoires and competitive exclusion. *Journal of Theoretical Biology* **169**, 375–390 (1994).

\[9\] Karaman, S. & Frazzoli, E. Incremental sampling-based algorithms for a class of pursuit-evasion games. *Vol. 68*, 71–87 (2010).

\[10\] Isaacs, R. *Differential Games I: Introduction* (RAND Corporation, 1954).

\[11\] Chung, T., Hollinger, G. & Isler, V. Search and pursuit-evasion in mobile robotics. *Auton. Robots* **31** (2011).

\[12\] Robin, C. & Lacroix, S. Multi-robot Target Detection and Tracking: Taxonomy and Survey. *Autonomous Robots* **40**, pp.729–760 (2016).

\[13\] Kumar, V., Rus, D. & Singh, S. Robot and sensor networks for first responders. *IEEE Pervasive Computing* **3**, 24–33 (2004).

\[14\] Dunbabin, M. & Marques, L. Robots for environmental monitoring: Significant advancements and applications. *RAM* **19**, 24–39 (2012).

\[15\] Hilal, A. R. An intelligent sensor management framework for pervasive surveillance (2013).

\[16\] Garcia, E., Casbeer, D. W., Pham, K. & Pachter, M. Cooperative aircraft defense from an attacking missile. In *CDC*, pp.2926–2931 (2014).

\[17\] Shishika, D. & Kumar, V. A review of multi agent perimeter defense games. In *Decision and Game Theory for Security*, pp.472–485 (2020).

\[18\] Shishika, D., Paulos, J. & Kumar, V. Cooperative team strategies for multi-layer perimeter-defense games. *RAL* **5**, 2738–2745 (2020).

\[19\] Von Moll, A., Shishika, D., Fuchs, Z. & Dorothy, M. The turret-runner-penetrator differential game. In *ACC*, pp.3202–3209 (2021).

\[20\] Lee, E. S., Shishika, D., Loiano, G. & Kumar, V. Defending a perimeter from a ground intruder using an aerial defender: Theory and practice. In *SSRR*, pp.184–189 (2021).
The optimal defender distribution $P_d^*$ is found by minimizing (3) with respect to $P_d$, subject to constraints of non-negativity ($P_d \geq 0$) and normalization ($\sum_d P_d = 1$). We can define an augmented Lagrangian as

$$L = \sum_a Q_a \tilde{F}_a + \lambda \left( \sum_d P_d - 1 \right) - \sum_d \nu_d P_d$$

where $\lambda$ and $\nu_d$ are Lagrange multipliers and set its derivative to zero to get the optimality conditions

$$\sum_a Q_a \tilde{F}_a^{\alpha} \Gamma(1+\alpha) = -\lambda^* + \nu_d^*$$
$$\nu_d^* \geq 0 \quad \text{(dual feasibility)}$$
$$\nu_d^* P_d^* = 0 \quad \text{(complementary slackness)}$$

If $P_d^* > 0$ for some $d$, we have $\nu_d^* = 0$ and thus $\sum_a Q_a \tilde{F}_a^{\alpha} \Gamma(1+\alpha) = -\lambda^*$. We can solve these equations for specific choices of cross-reactivity $f_{d,a}$.

If the shape-space is continuous and the cross-reactivity is $f_{d,a} = f(d-a)$, then we can write the first condition as

$$\int da \sum_a Q_a \tilde{F}(\tilde{P}_a^*) f(d-a) = -\lambda^*$$

where the total probability of tackling attacker $a$ is again a convolution $\tilde{P}_a^* = \int da \tilde{F}(\tilde{P}_d^*) f(d-a) = -c$ satisfies this equation for some constant $c$ (the convolution of a constant is a constant). Given such a $\tilde{P}_a^*$ we can calculate the optimal defender distribution as

$$P_d^* = F^{-1}\left[ \tilde{F}(\tilde{P}_a^*) F[f] \right]$$

where $F[\cdot]$ denotes the Fourier transform.