Quantal Response-Based Defense in One vs. Many Stackelberg Security Games

Gabriel Smithline

Tufts University, gabriel.smithline@tufts.edu

Abstract. This paper explores defensive strategies in 1 vs. many Stackelberg games with partial information. By utilizing the Quantal Response Model (QRM) and modeling its λ parameter as a stochastic process representing bounded rationality, I examine how myopic attackers influence each other and how a defender's bounded rationality affects strategy selection, while leveraging attacker congestion to the defender's benefit. The defender applies Bayesian learning to λ based on the potential of the attackers' strategy profiles, with λ modeled as a gamma distribution sampled to obtain its value. I empirically evaluate outcomes in games with varying upper and lower bounds for λ to analyze how different levels of bounded rationality impact the defender's performance and overall game outcomes. This study aims to enhance understanding of gameplay in partial information scenarios, using the Price of Anarchy (PoA) of the attacking agents as a metric for success.

Keywords: Security · Game Theory · Computer Science · Stackelberg Games · Deception · Quantal Response Model

1 Introduction, Related Work, and Contributions

Game Theory has its roots in economics, where John Nash established the concept of Nash Equilibrium, stating that in a non-cooperative game, players reach an equilibrium where no player can gain by unilaterally changing their strategy [9]. Since then, contributions from economics, mathematics, and computer science have expanded its applications across many domains. This paper focuses on security, utilizing concepts from Stackelberg Games, Quantal Response Model (QRM) [7], and Exact Potential games [8]. A Stackelberg game involves a leader (defender) setting their strategy first, followed by followers (attackers) who respond based on the leader's choice. Exact potential games, where the strategy one attacker plays affects the value of the strategy all other attackers play, are used to model the interactions between multiple attackers. The nested structure of these games, referred to as a metagame, offers a complex but insightful lens to analyze strategic interactions.

Applying game theory to security defense is well studied, traditionally assuming actors aim to maximize utility at each turn. However, more recent research explores deception, where actors do not play strictly optimally to deceive adversaries. This has primarily been studied from the attacker's perspective and

has been explored from the defender's perspective less so. Defensive deception can involve signaling strategies, honeypots, and information asymmetry. Notable contributions include [4], [3], and [11]. This paper builds on some foundations by applying the Quantal Response Model to defensive deception in Stackelberg Security games. There has also been work in congestion analysis, [12], [6], which can help put this work into context. Where a game theoretic approach is taken to understand the cost of these adversarial scenarios, and it's found that often times allowing congestion to build or using it as part of defense, can lead to a better outcome for the defender if they had not, as all other attackers are affected by the congestion and not nearly as successful. This helps to put us in context the real impacts of using deception and congestion as a defensive strategy in these partially observable games, where there is much information asymmetry.

The inspiration for this game structure comes from how often in cybersecurity scenarios attackers are not just playing against, say, a single defender, they are playing against all other attackers; each with their own preferences which are unknown to the single defender and all other attackers. Hence, attackers affect how other attackers play, and the defender needs to respond to the aggregate of the attackers; without knowledge of what each attacker's preferences over targets are. To derive these dynamics, the attackers compute their strategy profile by playing an iterated best response game, which halts either after 1000 iterations or when we reach an ϵ-Nash strategy profile for the attackers. I derive an exact potential function to capture the dynamics of this game; with this potential function, the defender aims to use it to capture the dynamics of all the attackers to compute his expected level of bounded rationality, or λ to play. Further inspiration also comes from work regarding network congestion and defensive strategies, where it's found in many cases in practice it can be cheaper and better to allow for network congestion when dealing with large amounts of attackers, as in aggregate they affect each other, and the cost to the system is nearly not as high as it could be.

I run various agent-based models (ABM) to simulate this; then analyze our results, focusing on the Price of Anarchy (PoA) [10].

1.1 Contributions

The key contributions of this paper are:

- Applying the Quantal Response Model (QRM) in the context of defensive deception, where a defender leverages attacker congestion to their benefit.
- Viewing bounded rationality as a stochastic process in Stackelberg security games, where from this I model and learn the level of bounded rationality to play in these games of partial information.
- Utilizing Bayesian learning to adaptively estimate the λ parameter, improving the defender's strategy against multiple attackers.
- Empirically evaluating the impact of different levels of bounded rationality on defender performance and overall game outcomes, enhancing the understanding of gameplay in partial information scenarios.

2 Methodology

2.1 Notation

I will describe the attacker strategy profile as σ^a (the mixed strategy profile the attackers have computed in their inner congestion game), and σ^{a*} as the optimal strategy profile if there was a benevolent dictator. U_i^a denotes the utility of attacker i, and U^d denotes the utility of the singular defender. I denote the lower bound for λ as λ_{lower} and denote the upper bound as λ_{upper}.

2.2 Game Model

I model our game as a Stackelberg security game with one defender and multiple attackers. The defender computes their strategy using the Quantal Response Model (QRM), which I denote as σ^d, then the attackers respond by playing their strategy profiles, which I denote as σ^a. Attackers compute their ϵ-Nash strategy profiles by playing a congestion game between themselves. Each target has a penalty (P_j), reward (R_j), and congestion cost (C_j), all randomly initialized in the range $[1, 10]$.

2.3 Attacker Strategy Computation

Attackers play against the defender and other attackers, considering their preference vectors. The utility function for attacker i on target j is:

$$U_{ij}^a = b_1(1 - \hat{x}_j) \cdot R_j^2 - b_2 \cdot \hat{x}_j \cdot P_j^2 - b_3(C_j \cdot N_j^2)$$

where $[b_1, b_2, b_3]$ is a randomly initialized normalized preference vector for each attacker. The optimization problem is:

$$\max \quad \mathbb{E}[U_i^a(x)] = \sum_{j=1}^{N} y_j \cdot U_{ij}^a$$

s.t.

$$\sum_{j=1}^{N} y_j = 1$$

$$0 \leq y_j \leq 1, \quad \forall j \in \{1, \ldots, N\}$$

where y represents the attacker's mixed strategy over N targets. Simply, each attacker is trying to maximize their utility with respect to their preferences, and their current view of the system.

2.4 Computation of ϵ-Nash Strategy Profile

To compute the ϵ-Nash strategy profile, I use the Iterated Best Response (IBR) algorithm:

Algorithm 1 Iterated Best Response Algorithm

while not at max_iterations and none_can_deviate **do**
  Randomly choose an attacker i from the attacking set
  Compute new strategy σ_i'
  if $U_i(\sigma_i') > U_i(\sigma_i) + \epsilon$ **then**
    Update strategy: $\sigma_i \leftarrow \sigma_i'$
    Move i to final check set
  else if $U_i(\sigma_i') > U_i(\sigma_i)$ **then**
    Update strategy: $\sigma_i \leftarrow \sigma_i'$
  else
    Keep current strategy: $\sigma_i \leftarrow \sigma_i$
    Move i to final check set
  end if
  if all attackers in final check set **then**
    for each attacker i in the final check set **do**
      if $U_i(\sigma_i') > U_i(\sigma_i) + \epsilon$ **then**
        Add i back to the attacking set
      end if
    end for
  end if
end while
return strategy profile σ^a once all attackers cannot deviate by more than ϵ or max_iterations is met

The result is σ^a, the ϵ-Nash strategy profile.

2.5 Exact Potential Function and Congestion Game

Attackers simultaneously play against all other attackers using a congestion model with an exact potential function:

$$\Phi(\sigma^a) = \sum_{j=1}^{n} \sum_{i=1}^{n} y_{ij} \cdot U_{ij}^{a}(\hat{x}_j, N_j) \tag{1}$$

This allows me to compute the potential of the strategy profile of all attackers.

Lemma 1. *The attackers' congestion subgame with utilities* $U_{ij}(\hat{x}_j, n_j) = (1 - \hat{x}_j)\, R_j - \hat{x}_j\, P_a^j - c_j\, n_j$ *and potential* $\Phi(y) = \sum_{j=1}^{t} \sum_{i=1}^{n} y_{ij}\, U_{ij}(\hat{x}_j, n_j)$ *is an* exact potential game*: for any player* i *and unilateral change from* y_i *to* y_i'*,* $\Delta U_i = \Delta \Phi$*.*

See Appendix A for the full proof of Lemma 1.

2.6 Defender Strategy Computation

The defender's utility function to maximize their mixed strategy profile σ^d is:

$$U_j^d = \hat{x}_j \cdot P_j - (1 - \hat{x}_j) \cdot R_j^2 + C_j \cdot \hat{N}_j^2$$

where $\hat{N}_j^2$ is the squared expected congestion.

Computation of $\boldsymbol{\lambda}$ The Quantal Response Value for attacker i at target j is:

$$\hat{Y}_{i,j} = \frac{e^{(\lambda \cdot U_{i,j}^a)}}{\sum_{k=1,k\neq i}^{n} e^{(\lambda \cdot U_{k,j}^a)}}$$

I model λ as a gamma random variable. The prior of λ is:

$$\text{Prior}(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

where α and β are shape and rate parameters. The likelihood of λ is:

$$L(\lambda \mid p) = \prod_{i=1}^{k} e^{(-\lambda p_i)} = e^{\left(-\lambda \sum_{i=1}^{k} p_i\right)}$$

The posterior is:

$$\text{Posterior}(\lambda \mid p) \propto L(\lambda \mid p) \times \text{Prior}(\lambda) \text{ for } \lambda_{\min} \leq \lambda \leq \lambda_{\max}$$

I compute $\mathbb{E}[\lambda]$ as:

$$\frac{\int_{\lambda_{\min}}^{\lambda_{\max}} \lambda \cdot \lambda^{\alpha-1} e^{\left(-\lambda(\beta+\sum_{i=1}^{k} p_i)\right) d\lambda}}{\int_{\lambda_{\min}}^{\lambda_{\max}} \lambda^{\alpha-1} e^{\left(-\lambda(\beta+\sum_{i=1}^{k} p_i)\right)} d\lambda}$$

This expected λ is used in the QRM computation. $\lambda_{\min}$ and $\lambda_{\max}$ are upper and lower bounds for λ in the interval I'm investigating at that point in the simulation.

Defender Strategy Computation Using the expected attacker strategy profiles, the defender solves the following optimization problem:

$$\max \mathbb{E}[U^d(\hat{y})] = \sum_{j=1}^{n} \left[x_j \cdot P_j^2 - (1 - x_j) \cdot R_j^2 + C_j \cdot (N_j)^2 \right]$$

s.t.

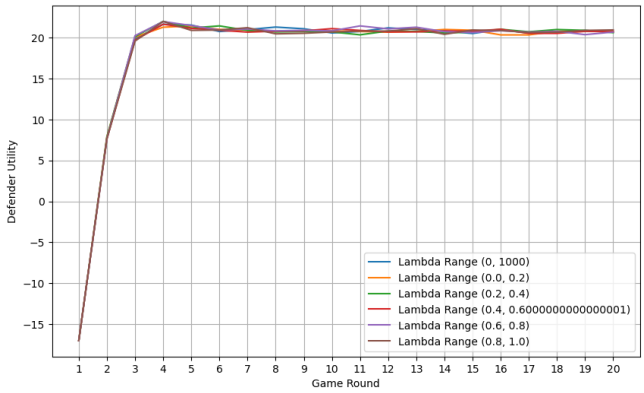
$$\sum_{j=1}^{N} x_j = 1$$

$$0 \leq x_j \leq 1, \quad \forall j \in \{1, \ldots, N\}$$

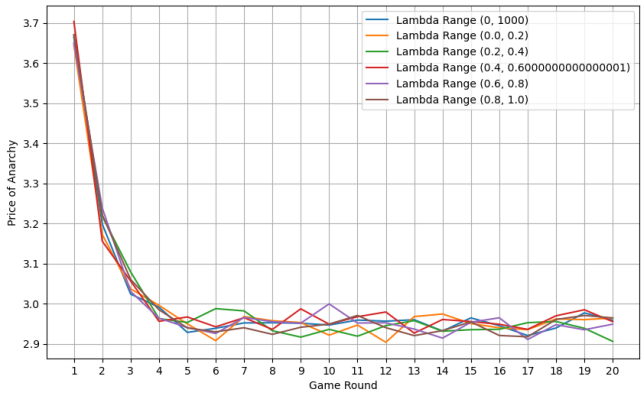
where x represents the defender's mixed strategy over N targets.

3 Experiments

I conducted various experiments to test the Quantal Response Model (QRM) in the 1 vs. many Stackelberg game format. Each game involved 5 attackers, 5 targets, and $\epsilon = 5$ for computing the ϵ-Nash strategy profile. Parameters P_j, R_j, and C_j were randomly assigned integers ranging from 1 to 10. I examined λ ranges from 0 to 1, divided into intervals (0 to 0.2, 0.2 to 0.4, etc.), and an unbounded range were λ ranged from 0 to 1000. The lower bound of the range is noted as $\lambda_{\min}$ and the upper bound as $\lambda_{\max}$. One key metric I use to track performance is the Price of Anarchy (PoA). PoA measures the efficiency of the attackers' strategy profile compared to the optimal strategy. Here, it is defined as the ratio of the potential of σ^{a*}, the optimal strategy profile for that attacker if I had a "benevolent dictator," to the potential of the computed strategy profile σ^a: PoA $= \frac{\Phi(\sigma^{a*})}{\Phi(\sigma^a)}$, where $\Phi(\sigma^{a*})$ is the potential of the optimal strategy profile, and $\Phi(\sigma^a)$ is the potential of the strategy profile computed by the attackers. I ran 600 games for each λ range, with each game consisting of 20 rounds (3,600 games for each seed, 10,800 games in total, and 216,000 game rounds were played in total).



(a) Trend of Defender Utility Over Game Rounds



(b) Trend of Price of Anarchy Over Game Rounds

Fig. 1: Comparison of Defender Utility and Price of Anarchy Trends

In figure **B** I see the trend of the PoA averaged across each lambda range averaged across all seeds. I see across the board our attackers are converged to a strategy profile with a relatively high PoA, around 3 to 2.9 across all λ ranges. In essence, at first, they do better, the PoA drops, but then eventually, over time they converge to a strategy profile that is significantly worse than what they could have played, about 3 times as worse. In figure **A**, I see our defender converging to a strategy profile that's better than where they started at the beginning of each game.

3.1 Analysis of Price of Anarchy Over Game Rounds

Looking at the key results for my agent-based simulations, a few things clearly jump out. This is a strong indication that in these 1 vs. Many Stackelberg games of partial information, the defender is able to play reasonably well against all the

attackers by using the QR Model and using potential functions as an aggregate to measure how to play against all attackers at once. I see that the attackers consistently converge to a strategy that's about 3 times worse than the best they could have done. What's also worth noting in my model is that there is no major difference in results when looking at the different λ ranges. This indicates that in these games of very partial information, the level of bounded rationality when making decisions can still be negligible in many cases, as in these security scenarios have very partial information; but still though if I play against the aggregate of all attackers, without knowing the specific preferences of attackers, or the specific strategy each attacker played rather just the impact that the total had, I can play reasonably well.

What's also worth noting is that there is a statistically significant positive correlation between λ and the defender utility in the lambda range of $0 - .2$, our lowest lambda range with the highest amount of bounded rationality. This is an indication that I need to induce high levels of bounded rationality in my model to see a clear-cut impact of modeling λ as a gamma distribution and applying learning to it.

3.2 Key Statistical Points

Table 1 summarizes the key statistical correlations observed in the experiments. These correlations provide insight into the relationship between the defender's utility and the Price of Anarchy (PoA) across different λ ranges.

Table 1: Key Statistical Correlations

λ **Range**	**Correlation (Defender Utility,** λ**)**	**Correlation (PoA,** λ**)**
$0 - 0.2$	0.81	-0.82
$0.2 - 0.4$	-0.21	0.29
$0.4 - 0.6$	-0.22	0.18
$0.6 - 0.8$	-0.21	0.30
$0.8 - 1.0$	-0.22	0.29
$0 - 1000$	0.055	0.13

When analyzing the features collected, it is evident that although the defender performed similarly across all λ ranges, there were some noteworthy points. In the λ range of 0 to 0.2, there was a 0.81 correlation between defender utility and λ value and a -0.82 correlation between PoA and λ value. This suggests that in scenarios with high bounded rationality, λ significantly impacts the defender's performance. As the defender becomes less bounded and more rational in this range, their performance improves, despite only knowing the aggregate response of the attackers.

In higher λ ranges, we observe different trends. The correlation between defender utility and λ becomes slightly negative, around -0.2. This indicates that

higher rationality does not always lead to better outcomes in partially observable environments for the defender. Instead, increased rationality may hinder the defender's performance, possibly due to overfitting to observed behaviors that are not fully representative of the attackers' strategies.

Conversely, the correlation between PoA and λ is slightly positive, ranging from 0.18 to 0.30, but these values do not show a strong increase as λ grows. This means that the increase in λ has a diminishing effect on the PoA, suggesting that beyond a certain level of rationality, further increases do not significantly impact attackers' efficiency.

In summary, these findings suggest that in highly uncertain and partially observable environments, incorporating bounded rationality and occasional randomization can enhance the defender's effectiveness. The significant positive correlation between λ and defender utility in the lowest λ range highlights the importance of stochastic elements in defensive strategies. However, as λ increases beyond a certain point, its impact on defender performance diminishes, indicating that playing more rationally does not always yield better results in these complex highly partially observable scenarios.

4 Conclusion and Future Work

In this work, I aimed to explore the augmented Quantal Response Model as a decision-making strategy in one vs. many Stackelberg security games. I explored how modeling different levels of bounded rationality can affect the mixed strategies played in these games, and through this, I examined how playing with more bounded rationality can affect the outcomes of the games of very partial information. I found through running many simulations that, λ although having a high correlation with defender utility at low bounded rationality levels, does not have much of an impact at higher levels and does not change much. But I also found, across the board, I was able to play strategies that induced the attackers to converge to non-optimal strategies relative to what they could have played.

4.1 Future Work

There are many directions one can take in experimenting here, as this was an initial exploration with some interesting results. The first direction identified is another subfield of game theory, called pursuer-evader games or differentiable game theory, for example [5], and [1]. Specifically, using Hamilton-Jacobi-Issac (HJI) equations to model this and develop much more rigorous strategies for affecting attackers [2] in a dynamical system setting. One other direction is now having smarter attacking agents; now they can work together, have memory, even be trained AI agents; can the Quantal Response Model hold up against these agents. Finally, applying more reinforcement learning or adaptive control techniques to explore deception is one area under investigation as well.

References

1. Buckdahn, R., Cardaliaguet, P., Quincampoix, M.: Some recent aspects of differential game theory. Dynamic Games and Applications **1**, 74–114 (2011)
2. Evans, L.C., Souganidis, P.E.: Differential games and representation formulas for solutions of hamilton-jacobi-isaacs equations. Indiana University mathematics journal **33**(5), 773–797 (1984)
3. Ferguson-Walter, K., Fugate, S., Mauger, J., Major, M.: Game theory for adaptive defensive cyber deception. In: Proceedings of the 6th Annual Symposium on Hot Topics in the Science of Security. HotSoS '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3314058.3314063, https://doi.org/10.1145/3314058.3314063
4. Fugate, S., Ferguson-Walter, K.: Artificial intelligence and game theory models for defending critical networks with cyber deception. AI Magazine **40**(1), 49–62 (Mar 2019). https://doi.org/10.1609/aimag.v40i1.2849, https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/2849
5. Garcia, E., Casbeer, D.W., Von Moll, A., Pachter, M.: Multiple pursuer multiple evader differential games. IEEE Transactions on Automatic Control **66**(5), 2345–2350 (2020)
6. Hausken, K., Zhuang, J.: Game theoretic analysis of congestion, safety and security. Traffic and Transportation Theory, Springer (2015)
7. McKelvey, R.D., Palfrey, T.R.: Quantal response equilibria for normal form games. Games and Economic Behavior **10**(1), 6–38 (1995). https://doi.org/https://doi.org/10.1006/game.1995.1023, https://www.sciencedirect.com/science/article/pii/S0899825685710238
8. Monderer, D., Shapley, L.S.: Potential games. Games and Economic Behavior **14**(1), 124–143 (1996). https://doi.org/https://doi.org/10.1006/game.1996.0044, https://www.sciencedirect.com/science/article/pii/S0899825696900445
9. Nash Jr, J.F.: Equilibrium points in n-person games. Proceedings of the national academy of sciences **36**(1), 48–49 (1950)
10. Roughgarden, T.: Selfish routing and the price of anarchy. MIT press (2005)
11. Sayed, M.A., Anwar, A., Kiekintveld, C., Kamhoua, C.: Honeypot allocation for cyber deception in dynamic tactical networks: A game theoretic approach (08 2023). https://doi.org/10.48550/arXiv.2308.11817
12. Zhang, M., Wu, J., Li, J., Reiher, P.: A game theoretical analysis of distributed denial-of-service defense incentive (2019)

A Proof of Lemma 1

A.1 Introduction

In congestion games, players choose resources, and their utility often depends on the number of other players choosing the same resources. We prove that the congestion game with the utility function below is a potential game by showing that the change in an individual player's utility from changing their strategy corresponds to the change in a global potential function.

A.2 Game Setup

Consider a game with n attackers (players) and t targets (resources). Let y_{ij} represent the strategy of player i attacking target j (a probability under mixed strategies), and let n_j be the number of players attacking target j.

A.3 Utility Function

The utility of attacker i for target j is:

$$U_{ij}(\hat{x}_j, n_j) = (1 - \hat{x}_j) \cdot R_j - \hat{x}_j \cdot P_a^j - c_j \cdot n_j. \tag{2}$$

The expected utility of attacker i under mixed strategy y_i is

$$U_i(y_i) = \sum_{j=1}^{t} y_{ij} \cdot U_{ij}(\hat{x}_j, n_j). \tag{3}$$

A.4 Potential Function

Define

$$\Phi(y) = \sum_{j=1}^{t} \sum_{i=1}^{n} y_{ij} \cdot U_{ij}(\hat{x}_j, n_j). \tag{4}$$

This function aggregates utilities across players and targets and internalizes congestion costs.

A.5 Proof of Lemma 1

A game is an exact potential game if there exists Φ such that, for any unilateral deviation of player i from y_i to y_i', the change in their utility equals the change in Φ.

Change in player i's utility. Let y and y' differ only in player i's strategy. Then

$$\begin{aligned}
\Delta U_i &= U_i(y_i') - U_i(y_i) \\
&= \sum_{j=1}^{t} \left[y_{ij}' \cdot U_{ij}(\hat{x}_j, n_j') - y_{ij} \cdot U_{ij}(\hat{x}_j, n_j) \right] \\
&= \sum_{j=1}^{t} (y_{ij}' - y_{ij}) \cdot \left[(1 - \hat{x}_j) \cdot R_j - \hat{x}_j \cdot P_a^j - c_j \cdot (n_j' - n_j) \right],
\end{aligned} \tag{5}$$

where n_j' is the number of attackers for target j under y'.

Change in the potential.

$$\begin{aligned}\Delta\Phi &= \Phi(y') - \Phi(y) \\ &= \sum_{j=1}^{t}\left(\sum_{k=1}^{n} y'_{kj}\cdot U_{kj}(\hat{x}_j, n'_j) - \sum_{k=1}^{n} y_{kj}\cdot U_{kj}(\hat{x}_j, n_j)\right) \\ &= \sum_{j=1}^{t}(y'_{ij} - y_{ij})\cdot\left[(1-\hat{x}_j)\cdot R_j - \hat{x}_j\cdot P_a^j - c_j\cdot(n'_j - n_j)\right], \end{aligned} \tag{6}$$

since the strategies of players $k \neq i$ do not change.

Equality. Comparing the two expressions gives

$$\Delta U_i = \Delta\Phi, \tag{7}$$

which establishes that the attackers' congestion subgame is an exact potential game.

Notes.

- Φ aggregates utilities over all players and targets (hence two summations), whereas U_i only aggregates attacker i's utilities.
- The equality $\Delta U_i = \Delta\Phi$ is the defining property of an exact potential game.