26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Strategic adversarial analysis of a game theory security model

Martin Wong

Defence Science and Technology Group, Email: martin.wong@defence.gov.au

Abstract: Stackelberg security games are models where a defender leads with a strategy that is observable to an attacker. Attackers then optimise their response against the observed defender strategy. By anticipating the attacker's optimal strategy, the defender can then lead with a strategy that will minimise the damage from the attack. This model formulation can be useful to inform how a defender should prepare for an unknown future attack when adopting a commitment-based strategy to mitigate a worst-case attack. This involves finding an optimal defensive strategy to a rational attacker strategy that result in a Stackelberg equilibrium strategy pair. This work describes an abstract security network model and the use of reinforcement learning to find the Stackelberg equilibrium defender/attacker pair to address the large action state spaces involved.

Keywords: *Stackelberg, security games, reinforcement learning, attack-defend model, network, game theory, modelling and simulation*

1. AN ABSTRACT NETWORK SECURITY MODEL

Our interest in Stackelberg games lies in their potential applicability to a range of graph-based adversarial scenarios. As such, the notion of an abstract attacker-defender model over a connected graph is first introduced with the key ideas being 1) that nodes in the graph have notional value, 2) that value is only accessible if it is reachable via paths in the graph and, 3) that agents have limited resources when taking actions in the environment. While specific applications are not discussed, we speculate that this base model (with the appropriate extensions) may fit scenarios focussed on cyber-physical attacks on critical infrastructure networks, containment of influence in social or information warfare networks and disruptions in mission dependency graphs or cyber-attack graphs.

Let $G = (V, E)$ be a connected graph, V is a set of n nodes and E is a set of m links where a node $v \in V$ represents an asset owned by a defender and produces some inherent utility u. The defender is represented by a special node d connected to the graph, which receives a utility flow $u(v)$ from node v if there exists a connected path $P(v, d)$ connecting v to d. The defender plays the role of the leader, making the first move and deciding a set of nodes $S_d \in G$ to reinforce during their turn within a resource budget. After the defender has committed to their strategy, the attacker observes the state of G and optimises their attack response S_a to the defender's S_d. An attacked node (if not protected) will block any utility flows that go through that node from reaching the defender node.

The defender receives an overall payoff $U_d = \sum_{k=0}^{n} u(v_k)$ over all connected paths $P(v_k, d)$. Similarly, the attacker receives a payoff $U_a = -U_d$ for its response strategy seeking to minimise the network utility reachable to the defender. The attacker's optimal response S_a^* would then be to maximise U_a for any given defender's S_d, where $S_a^*(S_d) = argmax_{S_a} U_a(S_d, S_a)$. In a Stackelberg game, the defender's goal would be to commit to a strategy S_d^* upfront such that it minimises the impact that the attacker causes through its optimal strategy S_a^*, that is $S_d^*(S_d) = argmax_{S_d} U_d(S_d, S_a^*(S_d))$. The Stackelberg equilibrium pair is then the tuple (S_d^*,S_a^*).

2. STACKELBERG ANALYSIS VIA REINFORCEMENT LEARNING

Using reinforcement learning to find the Stackelberg equilibrium [1-3], an alternating learning approach was employed to construct the policy models M_a and M_d that approximate S_a^* and S_d^*. An initial attacker policy model M_a^0 is first trained with defender strategies generated from an initial defender policy model M_d^{-1}. This attack model shall be the first iteration of S_a^*. Next, the defender policy model M_d^0 is trained against the recently learnt attacker model M_a^0 to approximate S_d^*. The learning step is then alternated between M_a and M_d at each training iteration. As M^i_a starts to approximate the optimal S_a^* given sufficient exposure to defender strategies generated by the previous M^{i-1}_d, then M^i_d should also approximate S_d^* as it trains against an improving model M_a^i. The strategy pair (S_d^i, S_a^i) generated by M_d^i and M_a^i is observed after each training iteration until the algorithm produces a strategy pair that remains unchanged with subsequent iterations[1]. At this point, the algorithm should have converged to a

[1] Assumes that the interactions between the defender and attacker strategies stabilises to a converged pair.

potential Stackelberg equilibrium pair[2]. Using the stable-baselines3 [4] version of the actor-critic (A2C) algorithm to train the policy models M_a and M_d, Figure 1 shows an example output of running the A2C algorithm on a randomly generated thirty-node network.

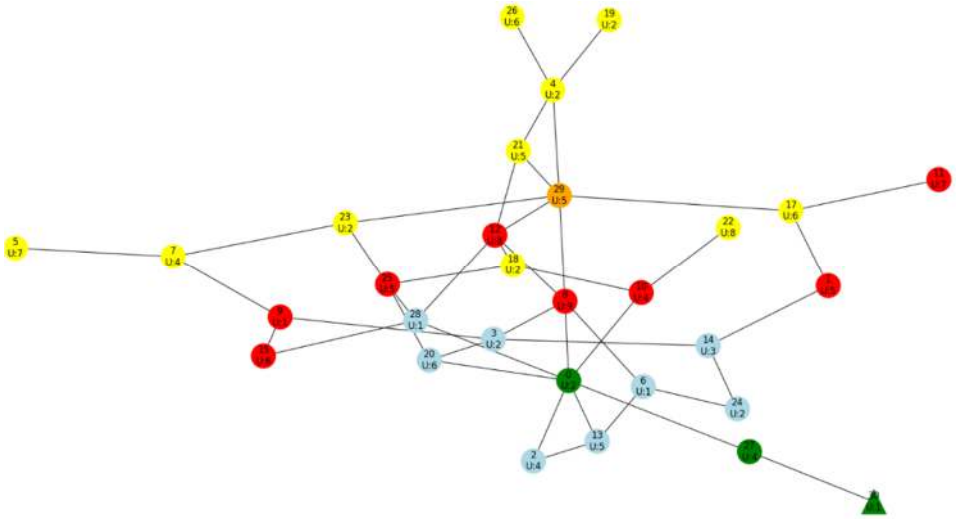

Figure 1: Defender's (green) versus attacker's (red) strategy. The defender learnt to secure mandatory nodes 0, 27, 30 (green nodes) as an attack on any of these nodes would cause the network to have no viable connected paths to the defender node 30 (triangle bottom right). The attacker was able to find an attack strategy that disabled (yellow and orange) a large segment of the network. Operational nodes are coloured light blue.

3. FUTURE WORK

The focus of this work is the development of an abstract network model to facilitate Stackelberg analysis, and the examination of reinforcement learning approaches as a viable and scalable solution for determining the equilibrium pair between two competing agents. The network security model provides a good base to consider further extensions to suit specific applications, such as nodes providing sets of multiple utility values or directionality of flow in the links. Modifications to player capabilities could incorporate aspects such as partial visibilities to the network, probabilistic outcomes of defender/attacker interaction outcomes and link based actions. We also plan to introduce an extension to the base model whereby the defender restores parts of the network to see how this would affect the Stackelberg equilibrium.

Reinforcement learning was utilised to find the Stackelberg equilibrium pair solutions via an alternating agent learning approach. While the approach seems initially promising, reaching equilibrium conditions reliably is still an ongoing issue as it depends on the choice of initial training parameters and the defined equilibrium condition. Further investigations are required to see whether the approach can effectively scale to increasing network sizes and future extensions introduced to the base model.

4. REFERENCES

[1] R. K. Mishra, D. Vasal and S. Vishwanath, "Model-free Reinforcement Learning for Stochastic Stackelberg Security Games," in *59th IEEE Conference on Decision and Control (CDC)*, Jeju, South Korea, 2020.

[2] D. Goel, A. Neumann, F. Neumann, H. Nguyen and M. Guo, "Evolving Reinforcement Learning Environment to Minimize Learner's Achievable Reward: An Application on Hardening Active Directory Systems," in *GECCO '23: Proceedings of the Genetic and Evolutionary Computation Conference*, 2023.

[3] G. Alcantara-Jimenez and J. B. Clempner, "Repeated Stackelberg security games: Learning with incomplete state information," *Reliability Engineering & System Safety,* vol. 195, no. 106695, 2020.

[4] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus and N. Dormann, "Stable-Baselines3: Reliable Reinforcement Learning Implementations," *Journal of Machine Learning Research,* vol. 22, no. 268, pp. 1-8, 2021.

[2] There may be more than one Stackelberg equilibrium pair.

