RL for Tower Defense with Evolutionary Towers Progress Report 1

Group 12
Andrew Wallace - 101210291 - andrewwallace3@cmail.carleton.ca
Mohammad Rehman - 101220514 - mohammadrehman@cmail.carleton.ca
Manal Hassan - 101263813 - manalhassa@cmail.carleton.ca
Derrick Zhang - 101232374 - derrickzhang@cmail.carleton.ca

MDP Specification

Environment Setup

The game takes place on a 10 by 10 grid. There is a fixed path that goes from the top left to the bottom right in an S-shape, covering about 15 cells. The remaining 85 cells are where towers can be placed.

The game has 10 waves total. The agent starts with 100 gold and the base has 20 lives. Each wave can last up to 200 time steps before moving to the next wave.

State Space

The state includes everything the agent needs to know at any moment:

Basic information:

- Current wave number (1 to 10)
- Current budget (starts at 100 gold, increases with kills)
- Base health (starts at 20 lives)
- Current time step within the wave (0 to 200)

Grid information:

For each cell in the 10 by 10 grid, we track:

- What is in the cell: empty, path, single-target tower, or area-of-effect tower
- If there is a tower, its level (1 to 5)
- If there is a tower, its experience points toward the next level

Enemy information:

For each enemy currently in the game (maximum 20 at once):

- Where it is on the path (position 0 to 14)
- How much health it has left

For the algorithm, we represent the state by combining the grid information (10 by 10 cells, with 3 values per cell: what is in it, tower level, and tower XP), enemy positions and health, and the scalar values into a single vector.

Action Space

To keep the problem manageable, the agent acts once at the beginning of each wave before enemies spawn. The agent can:

- Do nothing and save money
- Place a single-target tower on any empty non-path cell (costs 50 gold)
- Place an area-of-effect tower on any empty non-path cell (costs 80 gold)

This gives a total of 171 possible actions: 1 do-nothing action plus 2 tower types times 85 placeable positions. If a cell is occupied or the agent does not have enough money, those actions are hidden. This prevents the agent from wasting time trying invalid moves.

Transition Dynamics

After the agent places towers at the start of a wave, the game runs automatically:

Enemy spawning:

- Each wave spawns 5 + 2 (times the wave number enemies)
- Each enemy has 10 + (5 times the wave number) health
- Enemies appear gradually, one every 10 time steps
- All enemies start at position 0 on the path

During each time step:

Enemies move first:

- Each enemy moves forward one cell every 2 time steps
- If an enemy reaches the end of the path, the base loses 1 life and that enemy is removed

Then towers attack:

Single-target towers:

- Can shoot enemies within 2 cells
- Target an enemy by some heuristic (closest to reaching the base perhaps)
- Attack once per time step
- Deal 10 times their level in damage

Area-of-effect towers:

- Can shoot at spots within 2 cells

- Hit all enemies within 2 cells of that spot
- Attack once every 2 time steps
- Deal 6 times their level in damage to each enemy hit

When an enemy is defeated:

- The agent gains 10 gold
- The tower that killed it gains 1 experience point
- When a tower gets enough experience points, it levels up

Tower evolution happens when a tower gets enough kills:

- Level 1 to 2: 10 kills
- Level 2 to 3: 20 more kills (30 total)
- Level 3 to 4: 30 more kills (60 total)
- Level 4 to 5: 40 more kills (100 total)
- Maximum level is 5

Wave and episode ending:

A wave ends when all enemies are defeated or have reached the base. Then agent gets to place towers again.

The episode ends in success if all 10 waves are completed and the base still has health remaining. The episode ends in failure if the base health reaches 0. We also stop the episode if it takes more than 2000 total time steps to prevent it from running forever.

Reward Function

The agent earns points based on its performance:

During gameplay:

- +10 for each enemy defeated
- -50 for each enemy that reaches the base
- +5 each time a tower levels up

After completing a wave:

- +20 if the wave was cleared without losing any lives
- +10 if the wave was cleared but some lives were lost

At the end of the episode:

- +200 for defeating all 10 waves
- -100 if the base is destroyed

We do not penalize the agent for placing towers because we want it to learn that placing towers is necessary. Later we can add efficiency rewards if needed.

The agent's goal is to maximize the total points over the entire episode. We use a discount factor of around 0.99, which means future rewards are worth slightly less than immediate rewards, but long-term planning is still very important.

Rationale

The agent only acts at the start of each wave rather than every time step. This reduces the number of decisions needed and lets the agent focus on strategic tower placement rather than micromanagement.

We use a fixed path so the environment is deterministic and the agent can learn consistent strategies. We can add path variation later if the basic version works well.

The enemy and tower numbers are balanced so early waves are easy to learn from, middle waves require good strategy, and late waves are very challenging and require optimized tower placement and evolution.

To validate our approach quickly, we may start with a simplified version:

- 5 by 5 grid instead of 10 by 10
- 3 waves instead of 10
- Only single-target towers (no area-of-effect)
- No tower evolution (all towers stay level 1)
- 3 enemies per wave with 20 health each

Once we get one algorithm working on this simple version, we will gradually add complexity:

- Add tower evolution
- Add area-of-effect towers
- Increase grid size to 10 by 10
- Increase to 10 waves

This incremental approach reduces risk and lets us identify problems early.