Assignment 5

Pranav Kasela 846965

Solution of the problem

The objective of the assignment is to maximize the reward in following condition:

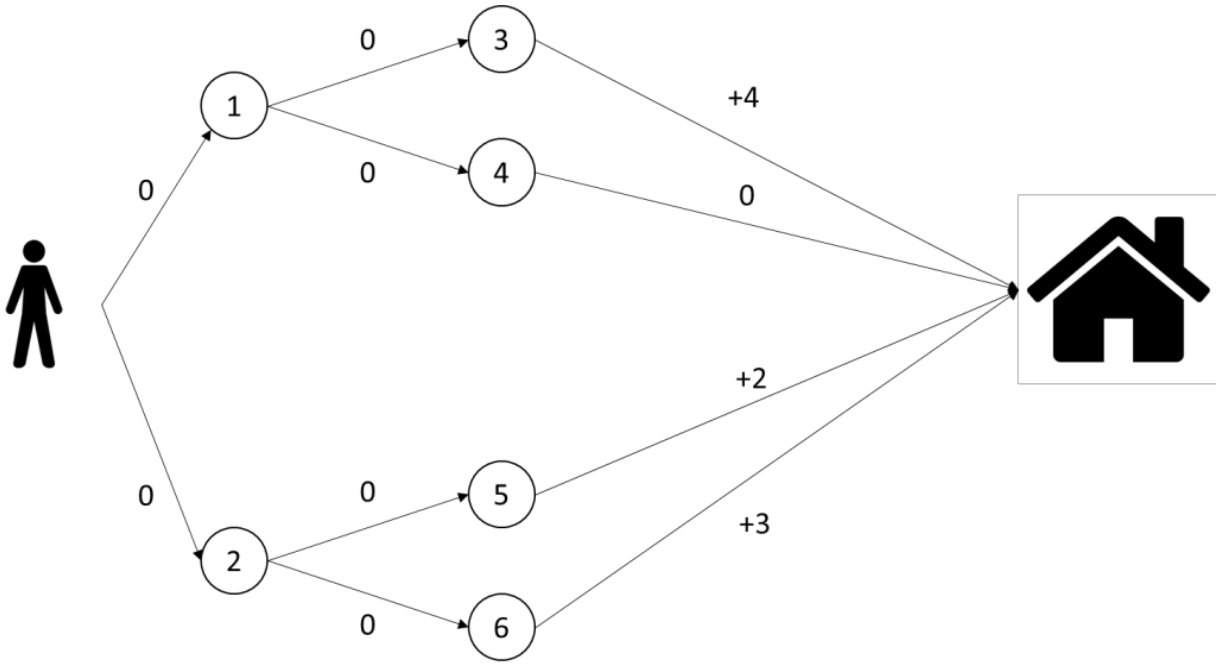

Figure 1: Representation of the problem

The initial polict π_0 is:

- $0 \to 2$
- $2 \to 5$
- $1 \to 4$

The state function of the nodes $3, 4, 5, 6$, which remains costant will be not be calculated or written.

The state function for the initial policy is:

$$V^{\pi_0}(0) = max\{0 + \gamma \cdot 0, 0 + \gamma \cdot 2\} = 2$$
$$V^{\pi_0}(1) = max\{0 + \gamma \cdot 4, 0 + \gamma \cdot 0\} = 4$$
$$V^{\pi_0}(2) = max\{0 + \gamma \cdot 2, 0 + \gamma \cdot 3\} = 3$$

where the first argument indicates the path to the left and the second argument indicates the path to the right. γ, the discount factor, has been decided to be $= 1$.

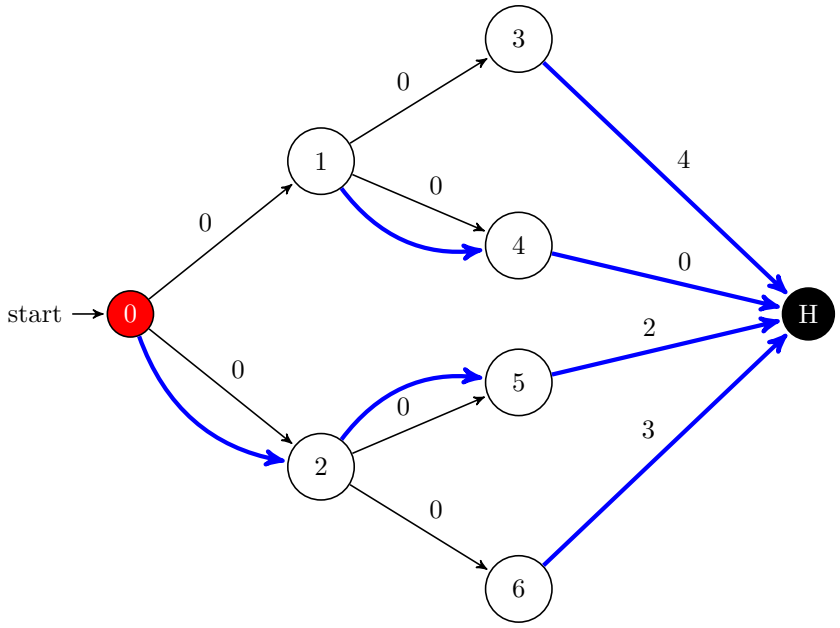

Figure 2: Inital Policy π_0 Representation

Policy π_1

The new policy π_1 now is:

- $0 \rightarrow 2$
- $2 \rightarrow 6$
- $1 \rightarrow 3$

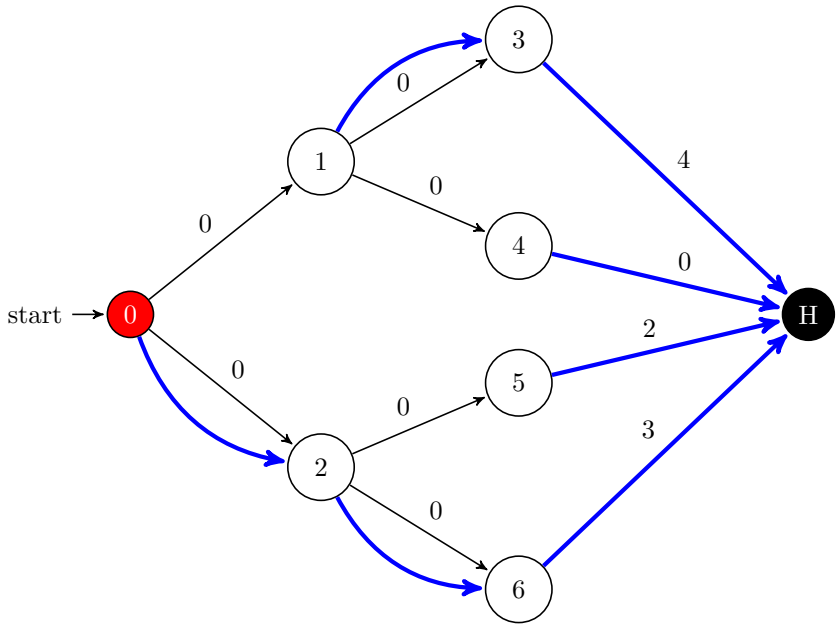

Figure 3: Policy π_1 Representation

The state function in π_1 is:

$$V^{\pi_1}(0) = max\{0 + \gamma \cdot 4, 0 + \gamma \cdot 3\} = 4$$
$$V^{\pi_1}(1) = max\{0 + \gamma \cdot 4, 0 + \gamma \cdot 0\} = 4$$
$$V^{\pi_1}(2) = max\{0 + \gamma \cdot 2, 0 + \gamma \cdot 3\} = 3$$

Policy π_2

The new policy derived from the state function is π_2:

- $0 \rightarrow 1$
- $2 \rightarrow 6$
- $1 \rightarrow 3$

Actually this is the best path, but to confirm the convergence the next step is needed.

It can be verified immediatly that the state function does not change and that $\pi_3 = \pi_2$, which is the condition for the convergence.

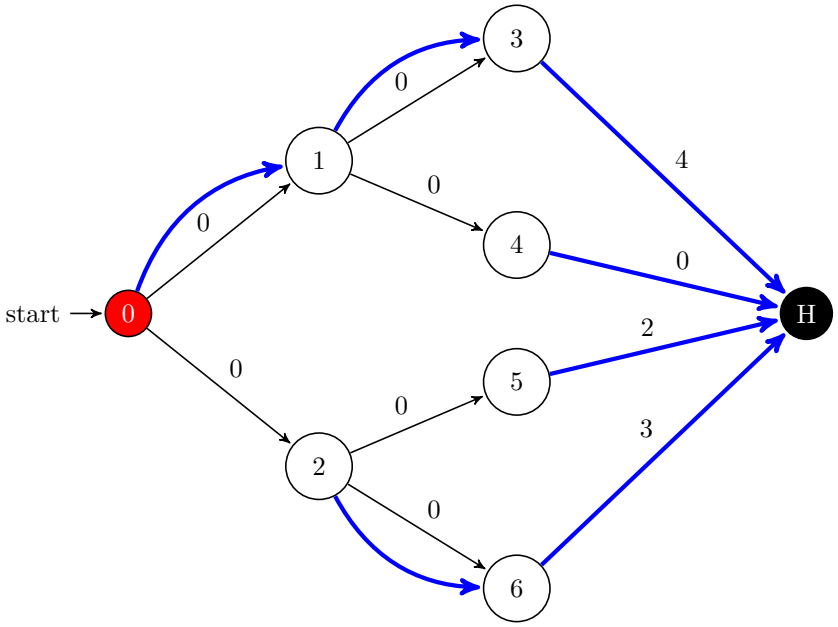

Figure 4: Policy $\pi_2 = \pi_3$ Representation

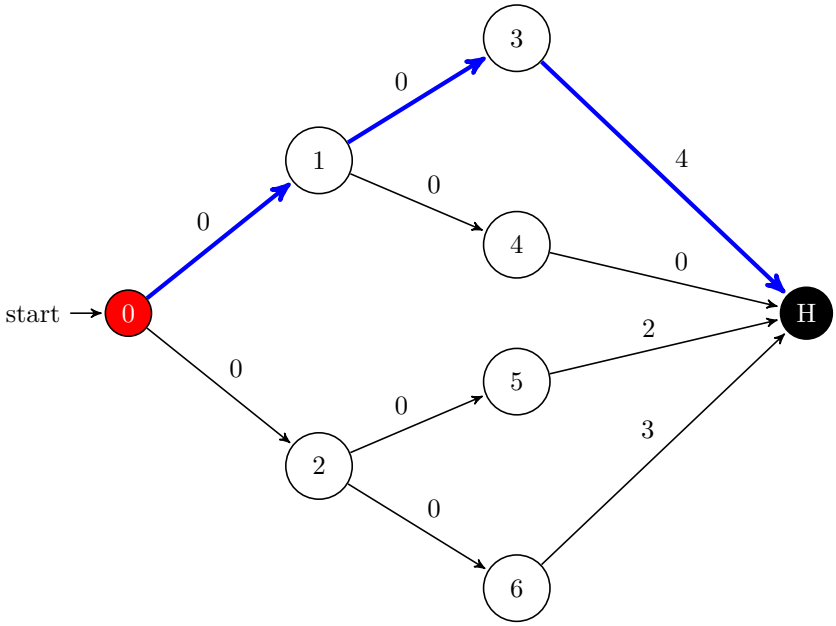

Figure 5: Final Solution Representation