Veeral Agarwal - 2019114009
P.Sahithi Reddy - 2020121011

Using roll number 2019114009 for all calculations.

SARSOP program is used to find the optimal policy from the POMDP formulated.
x = 1 - LastFourDigitsOfRollNumber)%30 + 1 / 100
x = 1 - 4009%30 + 1 / 100
x = 0.8, 1-x = 0.2
reward = 2019114009%90 + 10 = 19

Transition probability -
Success - 0.8
Failure - 0.2

There are a total of 128 states where each state is represented as a tuple Agent_Position>,Target_Position>,Call>)

Q1 : If you know the target is in (1,0) cell and your observation is o6 , what will be the initial belief state? Please submit the optimal policy file named RollNumber.policy for the POMDP taking into account the initial belief state you obtained.

Agent observes o6 observed when the target is not in the 1 cell neighborhood of the agent
Thus, the agent is out of the one cell neighbourhood of target i.e either of (0,1), (0,2), (0,3), (1,2), (1,3)

Each of the 5 states have equal probability and with the addition of 2 Calls, we get a total of 10 possible states. The initial belief state will be the 10 states having 0.1 probability each and the rest will be 0.

States	0	1	2	3
0	-	Agent	Agent	Agent
1	target	-	Agent	Agent

Q2 : If you are in (1,1) and you know the target is in your one neighborhood and is not making a call what is your initial belief state?

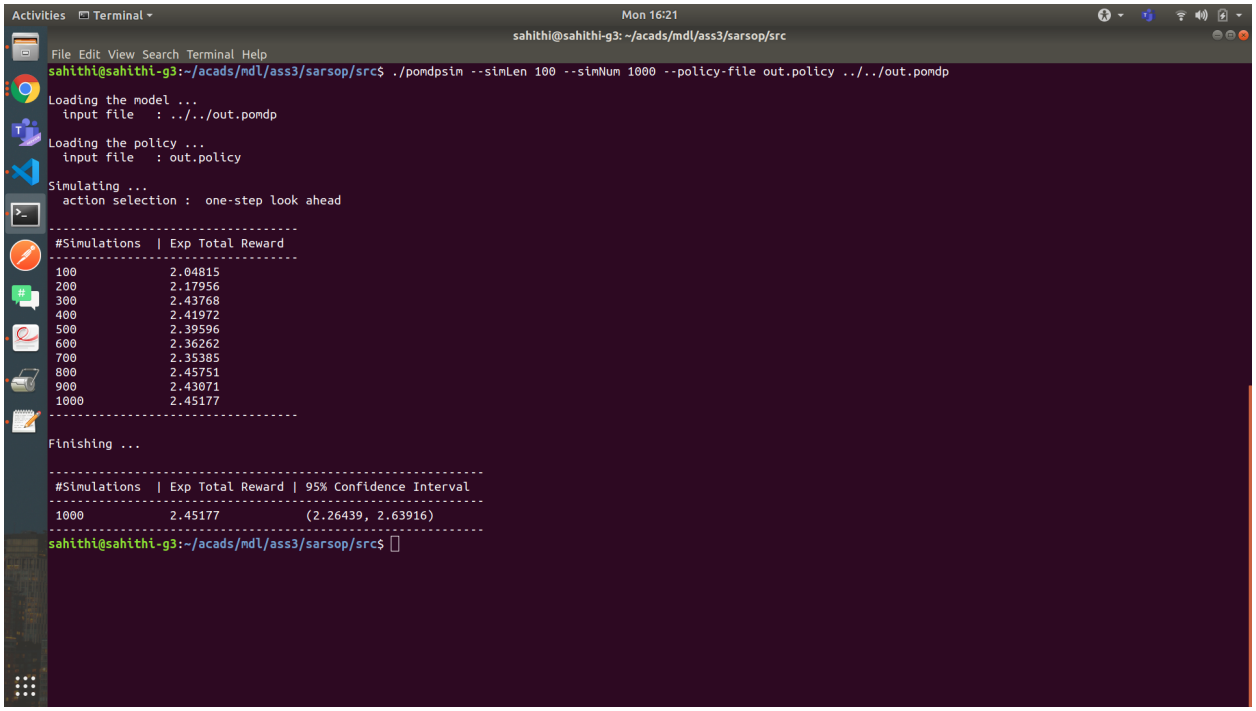
Given the agent is in (1,1) Target state is in one cell neighbourhood i.e either of (1,1), (0,1), (1,2), (1,0). Call is OFF Thus there are a total 4 states having an equal probability. The initial belief state will have these 4 states with probability 1/4 each and the rest will be 0.

states	0	1	2	3
0	-	Target	-	-
1	Target	Target,agent	Target	-

Q3 : What is the expected utility for initial belief states in questions 1 and 2?

With the pomdp calculated from Q1 and Q2, we use SARSOP to calculate the optimal policy and run simulations using that policy to calculate the expected rewards:

Expected reward for Q1 :
Running the simulations in steps of 100 for a total of 1000 times gave expected reward of **2.451**



Expected reward for Q2:

Running the simulations in steps of 100 for a total of 1000 times gave expected reward of **7.13**

```
sahithi@sahithi-g3:~/acads/mdl/ass3/sarsop/src$ ./pomdpsim --simLen 100 --simNum 1000 --policy-file out.policy ../../2019114009_b.pomdp

Loading the model ...
  input file   : ../../2019114009_b.pomdp

Loading the policy ...
  input file   : out.policy

Simulating ...
  action selection :  one-step look ahead

---------------------------------
 #Simulations  | Exp Total Reward
---------------------------------
 100            7.25675
 200            6.94897
 300            7.02134
 400            6.99666
 500            6.96871
 600            7.06334
 700            7.07649
 800            7.07116
 900            7.09476
 1000           7.13342
---------------------------------

Finishing ...

------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
------------------------------------------------------------
 1000           7.13342           (6.94573, 7.32111)
```

Q4: If your agent is in (0,0) with probability 0.4 and in (1,3) with probability 0.6 and the target is in (0,1), (0,2), (1,1), and (1,2) with equal probability, which observation are you most likely to observe? Explain.

Agent is in (0,0) with probability 0.4 and in (1,3) with probability 0.6 and the target is in (0,1), (0,2), (1,1) and (1,2) with equal probability
If agent in (0,0) :

O1	O2	O3	O4	O5	O6
0	0.25	0	0	0	0.75

Consider when Agent is in (1,3), then the only possible condition when Agents detects o4 is when Target is in (1,2), else it detects o6

O1	O2	O3	O4	O5	O6
0	0	0	0.25	0	0.75

Taking weighted mean of tables with their concerned probabilities we get Final Table = 0.4*Table1 + 0.6*Table2

O1	O2	O3	O4	O5	O6
0	0.1	0	0.15	0	0.75

most likely observation to be observed is o6 with a probability of 0.75

Q5：How many policy trees are obtained in the case of question 4, explain?

```
sahithi@sahithi-g3:~/acads/mdl/ass3/sarsop/src$ ./pomdpsol ../../2019114009_d.pomdp

Loading the model ...
  input file   : ../../2019114009_d.pomdp
  loading time : 0.02s

SARSOP initializing ...
  initialization time : 0.00s

-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0      0       0        1.31101    5.37471    4.0637      5        1
 0.01   11      51       4.03652    4.0656     0.029085    18       13
 0.01   17      100      4.05671    4.06318    0.00646978  46       25
 0.01   21      151      4.05941    4.06189    0.00247477  74       40
 0.02   25      200      4.06069    4.0617     0.00100511  91       50
 0.02   26      207      4.06088    4.06169    0.000808725 98       51
-------------------------------------------------------------------------------

SARSOP finishing ...
  target precision reached
  target precision  : 0.001000
  precision reached : 0.000809

-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0.02   26      207      4.06088    4.06169    0.000808725 94       51
-------------------------------------------------------------------------------

Writing out policy ...
  output file : out.policy
```

Policy trees calculation:
N = (|o|^t -1) / (|o| -1)
Number of trees = ||A||^N

Here A refers to the actions possible o is the possible observations and T is time horizon (the number of steps the agent takes)
We have 5 actions, 6 observations, and time horizon (#trial) generated from the pomdpsol is 26.

Number of nodes N = 6^26 - 1 / 6 - 1 = 3.4116346e+19

The number of policy trees generated by these nodes are
|A|^N = 5^(3.4116346e+19)

This is a very large number because as we increase the horizon, the number of nodes will not converge. So, if we increase our horizon size there will be exponentially increasing policy trees .