**Carissa Ying Geok Teng (A0205190R/E0425113)**

1a)

**State**

On(Start), Path(s0, s1) …

**Action**

**Travel(from, to)**

Precond: On(from) ^ ~Visited(to) ^ Path(from, to)

Effect: ~On(from) ^ On(to) ^ Visited(to)

**Transition**

P(s1|s0, Travel(s0, s1)) …

**Reward**

R(s) = -0.4

R(Visited(s0) ^ Visited(s1) ^ …) = 1

1b)

**State**

ItemsInStorage(n), NumOfOrders(o), ~Ordered, Backordered(u)

**Action**

**Order(n, m):** Order 1 time such that inventory = M

Precond: ItemsInStorage(n) ^ (n + m = M) ^ ~Ordered

Effect: ~ItemsInStorage(n) ^ ItemsInStorage(n + m) ^ Ordered

**Backorder(n, u, o):** Backorder up to B units

Precond: ItemsInStorage(n) ^ Backordered(u) ^ (n + o < N) ^ (u + o < B)

Effect: ~ItemsInStorage(n) ^ ~Backordered(u) ^ ItemsInStorage(n + o) ^ Backordered(u + o)

**FulfillOrder(n, o, f)**

Precond: ItemsInStorage(n) ^ NumOfOrders(o) ^ (n – f ≥ 0) ^ (o – f ≥ 0)

Effect: ~ItemsInStorage(n) ^ ~NumOfOrders(o) ^ ItemsInStorage(n – f) ^ NumOfOrders(o – f)

**NextDay(n, o, u):**

Precond: ItemsInStorage(n) ^ NumOfOrders(o) ^ Backordered(u) ^ (o == 0)

Effect: ~Ordered ^ ~Backordered(u)

**Transition**

P(New inventory and orders | Old inventory and orders, Action to order more or fulfil orders)

**Reward**

R(s, Order(n, m)) = - c

R(s, Backorder(n, u, o)) = -b \* o

R(s, FulfillOrder(n, o, f)) = f

R(s, NextDay(n, o, u)) = -n

1c)

**State**

Screen display where each screen pixel with one of its values from 0-127

**Action**

One of the 18 actions

**Transition**

P(display after an action is taken| current screen display, one of the 18 actions)

**Reward**

∆Score - ∆Time

2a)

**Policy**

π\*(s1) = a2

π\*(s2) = a1

**Value Function**

U\*(s1) = P(s1|s1, a2)R(s1|a2) + P(s2|s1, a2)R(s2|a­2) = 0.1(0) + 0.9(3) = 2.7

U\*(s2) = P(s1|s2, a1)R(s1|a1) + P(s2|s2, a1)R(s2|a­1) = 0(1) + 1(3) = 3

2b)

**Policy**

π\*(s1, t1) = a2

π\*(s2, t1) = a1

π\*(s1, t2) = a2

π\*(s2, t2) = a1

**Value Function**

U\*(s1) = P(s1|s1, a2)R(s1|a2) + P(s2|s1, a2)R(s2|a­2)

+ P(s1|s1, a2) P(s1|s1, a2)R(s1|a2) + P(s2|s1, a2) P(s1|s1, a2)R(s2|a2) (s1 on first move)

+ P(s1|s2, a1)P(s2|s1, a2)R(s1|a­2) + P(s2|s2, a1)P(s2|s1, a2)R(s2|a­2) (s2 on first move)

= 0.1(0) + 0.9(3) + 0.1(0.1)(0) + 0.9(0.1)(3) + 0(0.9)(0) + 1(0.9)(3) = 5.67

U\*(s2) = 3 + 3 = 6

2c)

**Policy**

π\*(s1) = a2

π\*(s2) = a1

**Value Function**

U\*(s2) = = 3 / 0.1 = 30

U\*(s1) = γP(s1|s1, a2)U\*(s1) + γP(s2|s2, a2)U\*(s2)

= 0.9(0.1)U\*(s1) + 0.9(0.9)(30)

0.91U\*(s1) = 24.3

U\*(s1) = 26.703