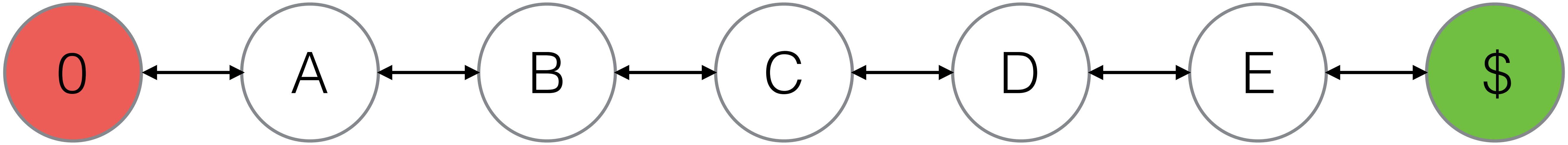
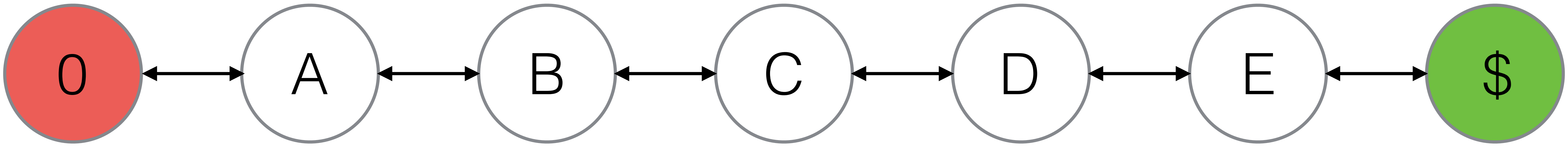
Random Walk Problem



Random Walk Problem



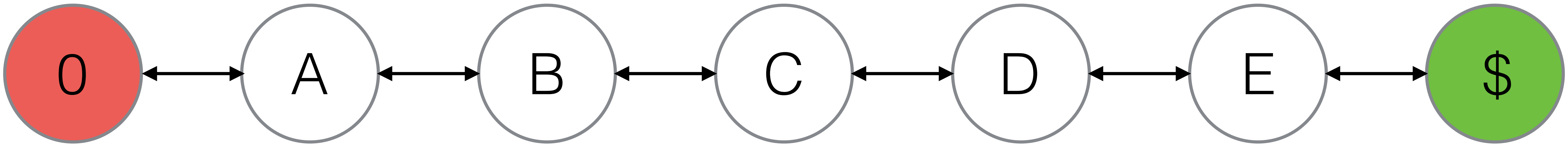
$$S := \{A, B, C, D, E\}$$

$$A := \{Left, Right\}$$

$$\phi := \{0, \$\}$$

$$R = \{-1, 0, 1\}$$

Random Walk Problem



$$S := \{A, B, C, D, E\}$$

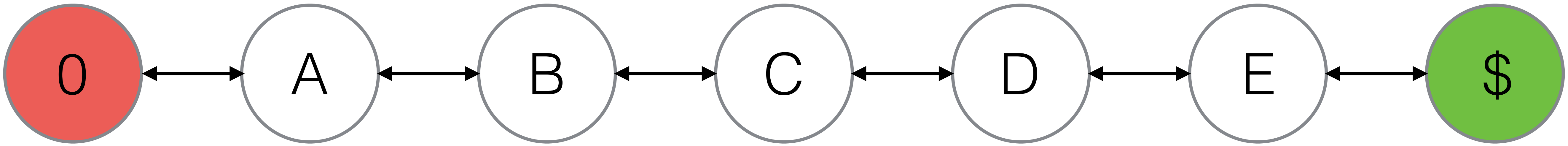
$$A := \{Left, Right\}$$

$$\phi := \{0, \$\}$$

$$R = \{-1, 0, 1\}$$

What about the transition probabilities?

Random Walk Problem



$$T = \infty$$

n = 2

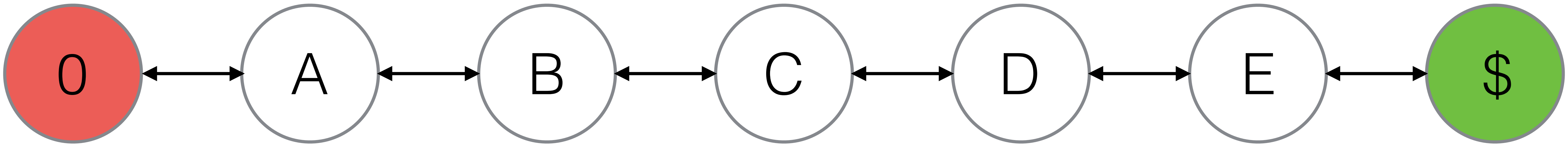
$$\pi = \epsilon - Greedy$$

St	C					
Rt	0					
At	r					

Initialize $Q(s, a)$ arbitrarily, $\forall s \in \mathcal{S}, a \in \mathcal{A}$
Initialize π to be ε-greedy with respect to Q, or to a fixed given policy
Parameters: step size $\alpha \in (0, 1]$, small $\varepsilon > 0$, a positive integer n
All store and access operations (for S_t, A_t, and R_t) can take their index mod n

Repeat (for each episode):
  Initialize and store $S_0 \neq$ terminal
  Select and store an action $A_0 \sim \pi(\cdot|S_0)$
  $T \leftarrow \infty$

Random Walk Problem



$T = \infty$

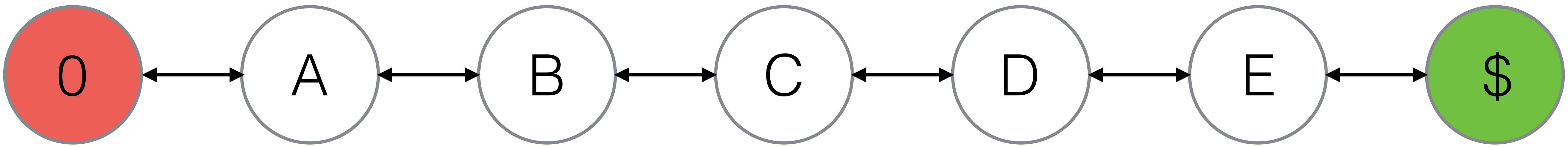
n = 2

t = 0

St	C	D				
Rt	0	0				
At	r	l				

For $t = 0, 1, 2, \ldots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$T = \infty$

n = 2

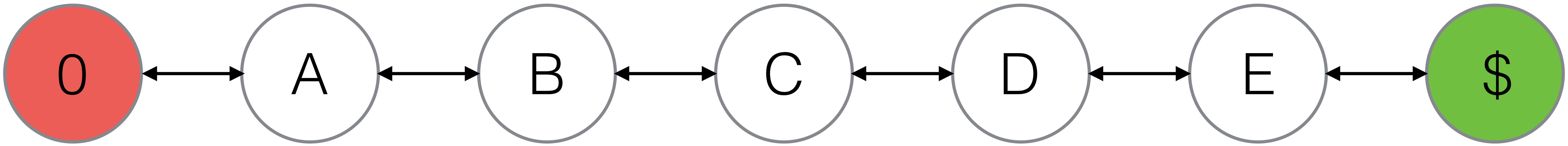
t = 0

tau = -1

St	C	D				
Rt	0	0				
At	r	l				

| $\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)
| If $\tau \geq 0$:
| $\quad G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$
| $\quad$ If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$
| $\quad Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha \left[G - Q(S_\tau, A_\tau)\right]$

Random Walk Problem



$T = \infty$

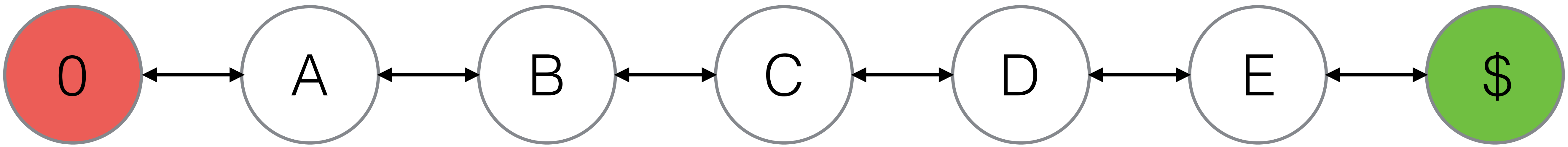
n = 2

t = 1

St	C	D	C			
Rt	0	0	0			
At	r	l	r			

For $t = 0, 1, 2, \ldots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$$T = \infty$$

n = 2

t = 1

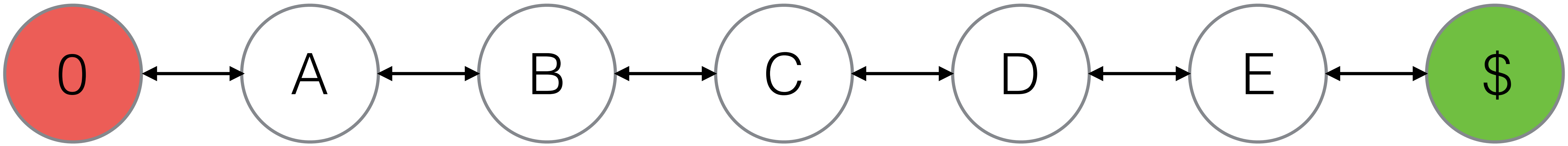
tau = 0

$$G = \gamma^0 R_1 + \gamma^1 R_2$$

St	C	D	C			
Rt	0	0	0			
At	r	l	r			

$\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)
If $\tau \geq 0$:
$G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$
If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$
$Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha\,[G - Q(S_\tau, A_\tau)]$

Random Walk Problem



$$T = \infty$$

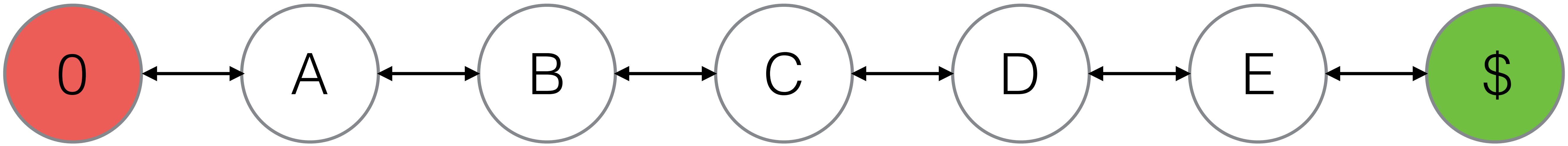
n = 2

t = 2

St	C	D	C	D		
Rt	0	0	0	0		
At	r	l	r	r		

For $t = 0, 1, 2, \ldots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$T = \infty$

n = 2

t = 2

tau = 1

St	C	D	C	D		
Rt	0	0	0	0		
At	r	l	r	r		

$\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)

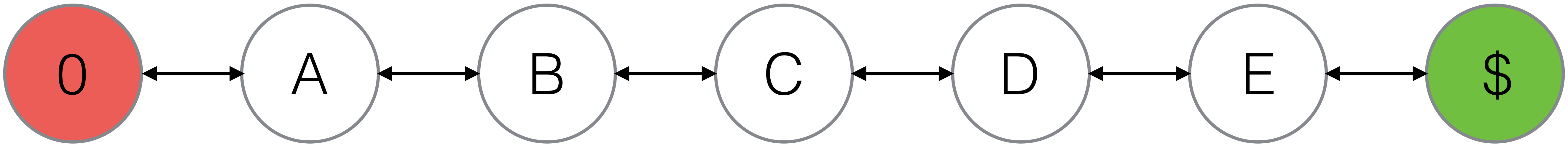
If $\tau \geq 0$:

$G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$

If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$

$Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha \left[G - Q(S_\tau, A_\tau)\right]$

Random Walk Problem



$T = \infty$

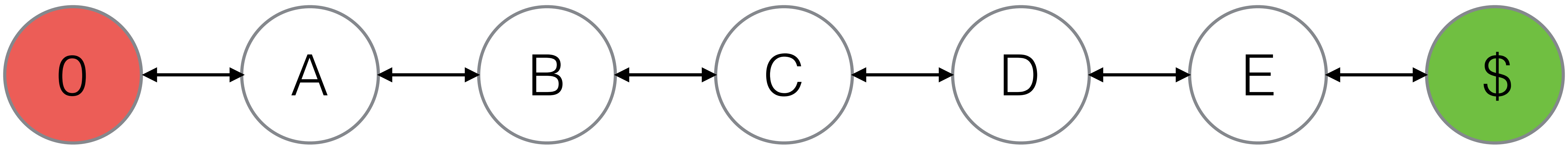
n = 2

t = 3

St	C	D	C	D	E	
Rt	0	0	0	0	0	
At	r	l	r	r	r	

For $t = 0, 1, 2, \ldots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot | S_{t+1})$

Random Walk Problem



$T = \infty$

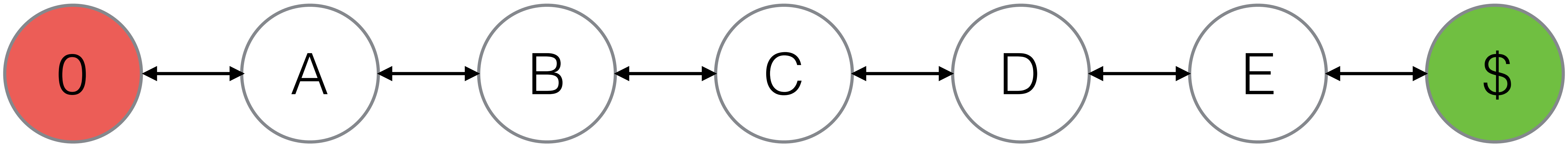
n = 2

t = 4

St	C	D	C	D	E	$
Rt	0	0	0	0	0	1
At	r	l	r	r	r	

For $t = 0, 1, 2, \ldots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$T = 5$

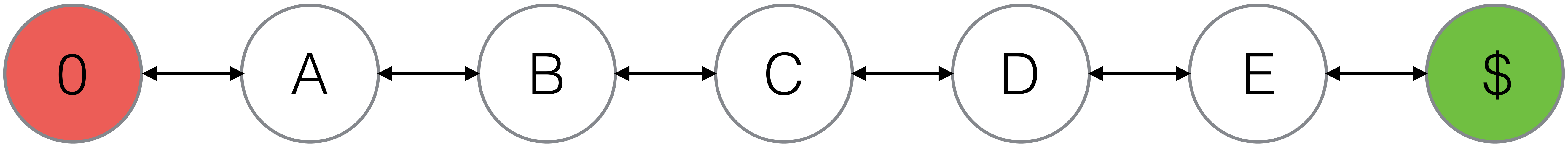
n = 2

t = 4

St	C	D	C	D	E	$
Rt	0	0	0	0	0	1
At	r	l	r	r	r	

For $t = 0, 1, 2, \dots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$$T = 5$$

n = 2

t = 4

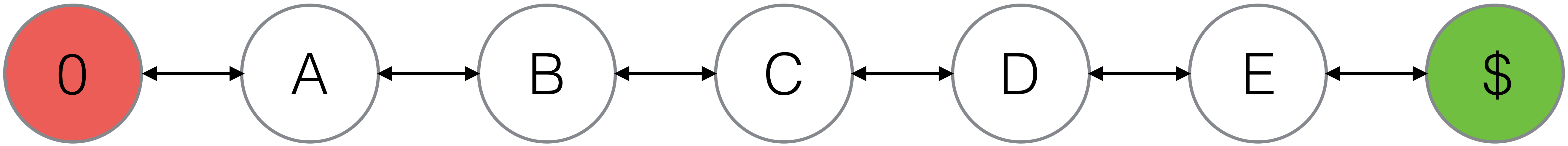
tau = 3

St	C	D	C	D	E	$
Rt	0	0	0	0	0	1
At	r	l	r	r	r	

$$G = \gamma^0 R_4 + \gamma^1 R_5 = 0 + 1$$
$$Q(D, r) = 0 + \alpha(1 - 0)$$

| $\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)
| If $\tau \geq 0$:
| $G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$
| If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$
| $Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha \left[G - Q(S_\tau, A_\tau)\right]$

Random Walk Problem



$T = 5$

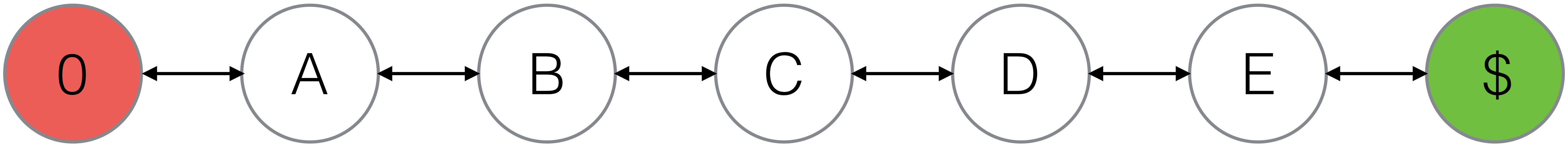
n = 2

t = 5

St	C	D	C	D	E	$
Rt	0	0	0	0	0	1
At	r	l	r	r	r	

For $t = 0, 1, 2, \ldots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$T = 5$

n = 2

t = 5

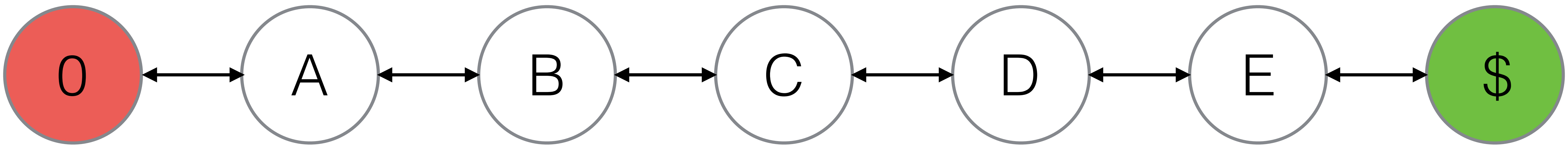
tau = 4

St	C	D	C	D	E	$
Rt	0	0	0	0	0	1
At	r	l	r	r	r	

$$G = \gamma^0 R_5$$
$$Q(E, r) = 0 + \alpha(1 - 0)$$

| $\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)
| If $\tau \geq 0$:
| $G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$
| If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$
| $Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha \left[G - Q(S_\tau, A_\tau)\right]$

Random Walk Problem



$T = 5$

n = 2

t = 5

tau = 4

St	C	D	C	D	E	$
Rt	0	0	0	0	0	1
At	r	l	r	r	r	

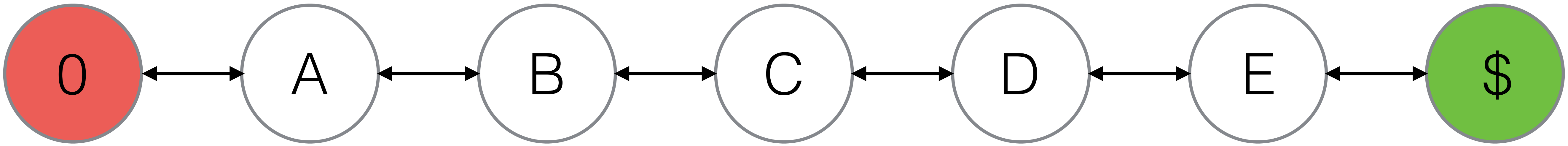
$$G = \gamma^0 R_5$$

$$Q(E, r) = 0 + \alpha(1 - 0)$$

| $\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)
| If $\tau \geq 0$:
| $G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$
| If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$
| $Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha \left[G - Q(S_\tau, A_\tau)\right]$

Until $\tau = T - 1$

Random Walk Problem



$T = \infty$

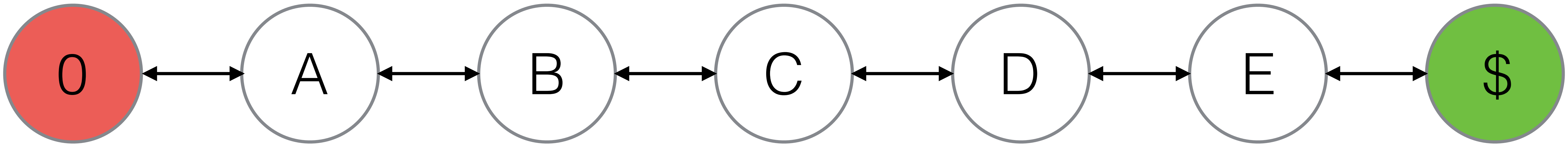
n = 2

St	C					
Rt	0					
At	r					

Initialize $Q(s, a)$ arbitrarily, $\forall s \in \mathcal{S}, a \in \mathcal{A}$
Initialize π to be ε-greedy with respect to Q, or to a fixed given policy
Parameters: step size $\alpha \in (0, 1]$, small $\varepsilon > 0$, a positive integer n
All store and access operations (for S_t, A_t, and R_t) can take their index mod n

Repeat (for each episode):
  Initialize and store $S_0 \neq$ terminal
  Select and store an action $A_0 \sim \pi(\cdot|S_0)$
  $T \leftarrow \infty$

Random Walk Problem



$T = \infty$

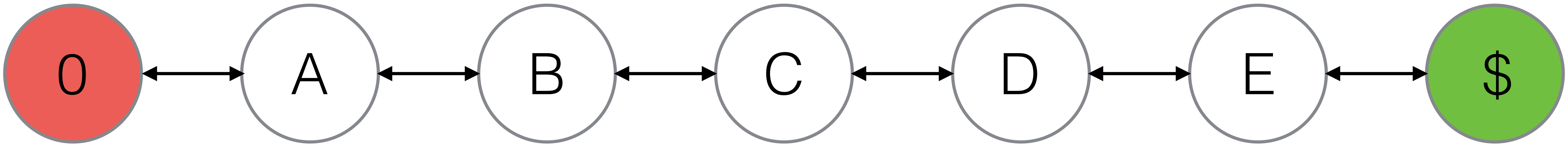
n = 2

t = 0

St	C	D				
Rt	0	0				
At	r	r				

For $t = 0, 1, 2, \dots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$$T = \infty$$

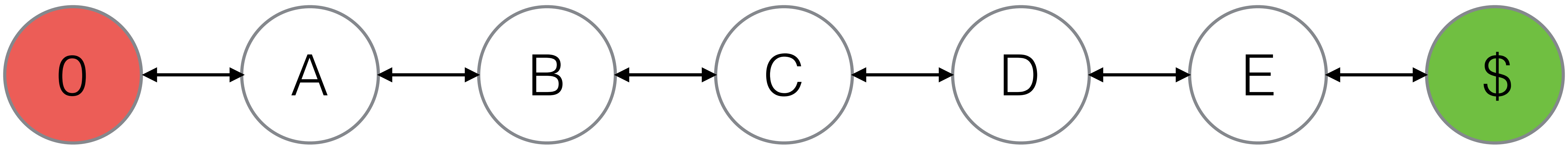
n = 2

t = 1

St	C	D	E			
Rt	0	0	0			
At	r	r	r			

For $t = 0, 1, 2, \ldots$:
| If $t < T$, then:
| Take action A_t
| Observe and store the next reward as R_{t+1} and the next state as S_{t+1}
| If S_{t+1} is terminal, then:
| $T \leftarrow t + 1$
| else:
| Select and store an action $A_{t+1} \sim \pi(\cdot|S_{t+1})$

Random Walk Problem



$T = \infty$

n = 2

t = 1

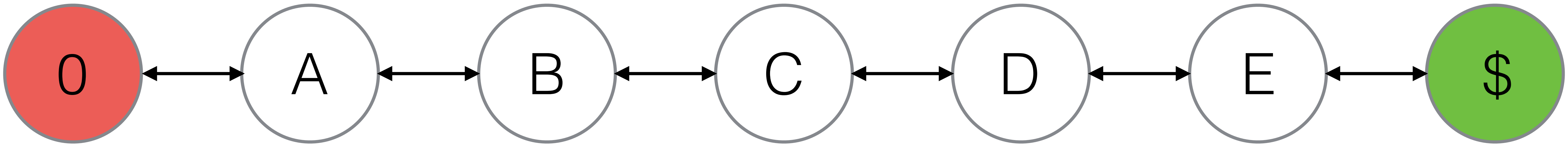
tau = 0

St	C	D	E			
Rt	0	0	0			
At	r	r	r			

$$G = \gamma^0 R_1 + \gamma^1 R_2$$

| $\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)
| If $\tau \geq 0$:
| $G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$
| If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$
| $Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha \left[G - Q(S_\tau, A_\tau)\right]$

Random Walk Problem



$T = \infty$

n = 2

t = 1

tau = 0

St	C	D	E			
Rt	0	0	0			
At	r	r	r			

$$G = \gamma^0 R_1 + \gamma^1 R_2$$

$$G = G + \gamma^2 Q(E, r)$$

| $\tau \leftarrow t - n + 1$ (τ is the time whose estimate is being updated)
| If $\tau \geq 0$:
| $G \leftarrow \sum_{i=\tau+1}^{\min(\tau+n,T)} \gamma^{i-\tau-1} R_i$
| If $\tau + n < T$, then $G \leftarrow G + \gamma^n Q(S_{\tau+n}, A_{\tau+n})$
| $Q(S_\tau, A_\tau) \leftarrow Q(S_\tau, A_\tau) + \alpha \left[G - Q(S_\tau, A_\tau)\right]$