

c_t

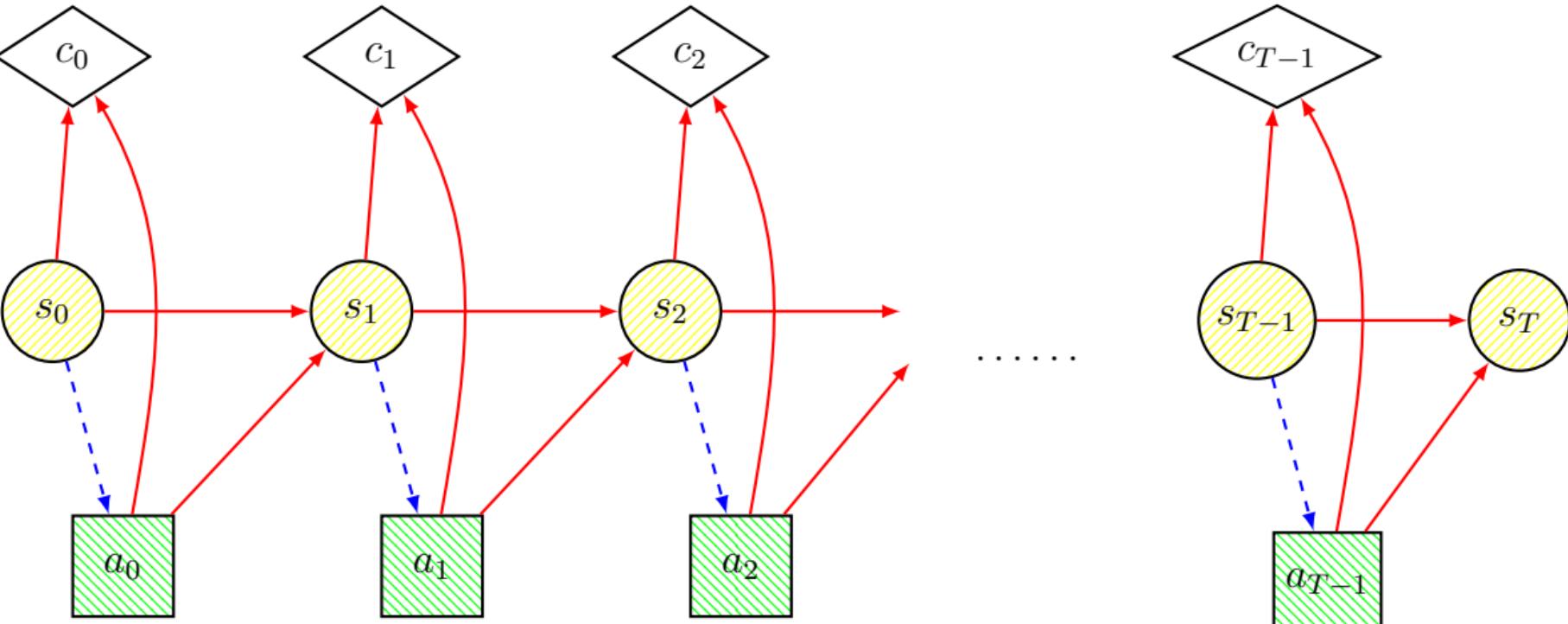
Cost or Loss
or Utility or Reward
at time t

s_t

State Random Variable
at time t

a_t

Action or Decision
or Policy taken
at time t



Policy Function could be modelled as a Neural Network $\pi_\theta(a_t|s_t)$



State Evolving Function $p(s_{t+1}|s_t, a_t)$ and Cost Function $c_t = c(s_t, a_t)$