Rewards Utility r_t or nagative Cost c_t

States $s_t \in \mathbb{S}$

observations of states $o_t \in \mathbb{O}$

Agent(Bridge Manager)

Environment(Bridge In Operation)

Taking actions $a_t \in \mathbb{A}$