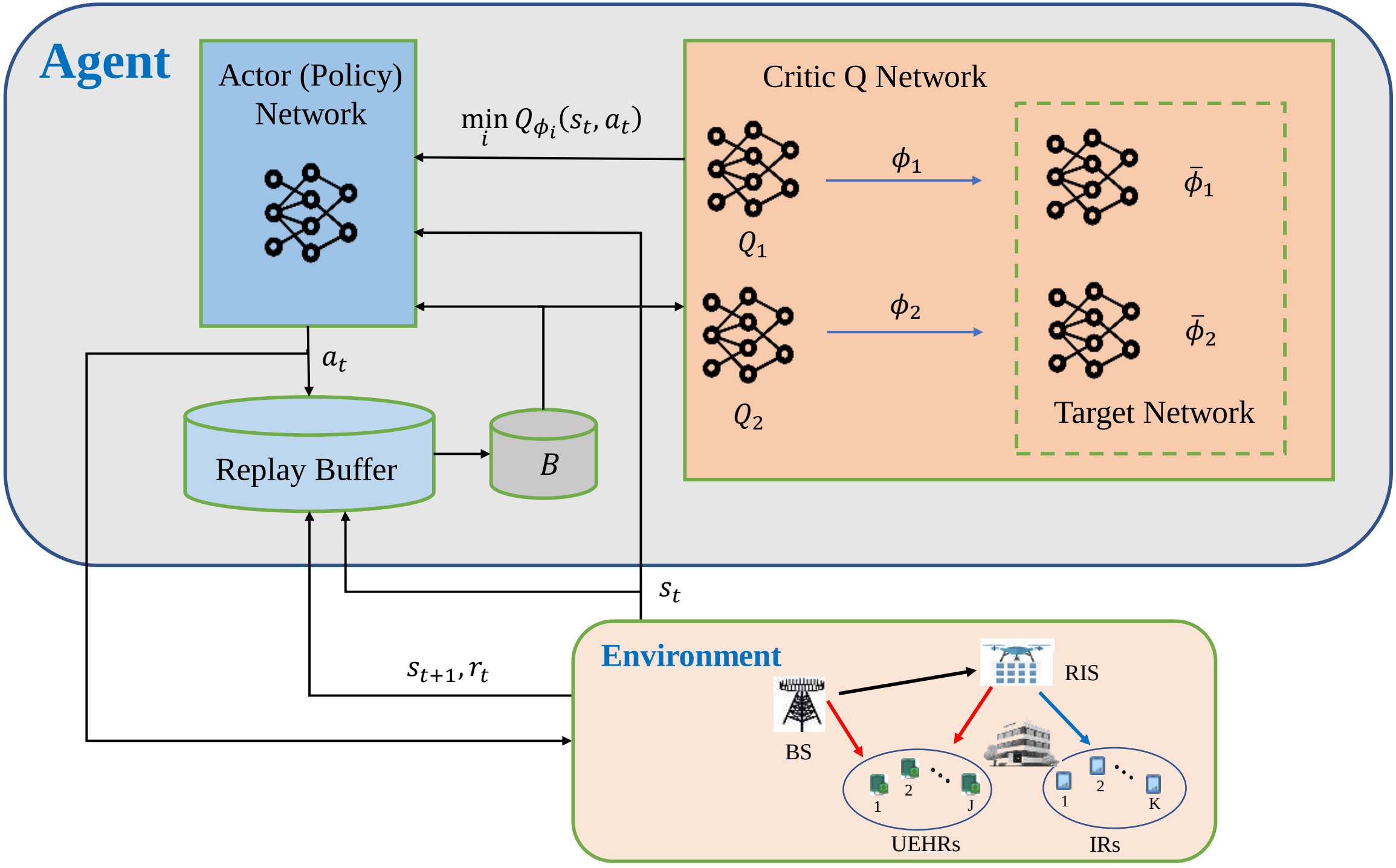

Agent
Actor (Policy) Network
$\min_i Q_{\phi_i}(s_t, a_t)$
Critic Q Network
ϕ_1
$\bar{\phi}_1$
Q_1
ϕ_2
$\bar{\phi}_2$
Q_2
Target Network
a_t
Replay Buffer
B
s_t
s_{t+1}, r_t
Environment
BS
RIS
1
2
J
UEHRs
1
2
K
IRs