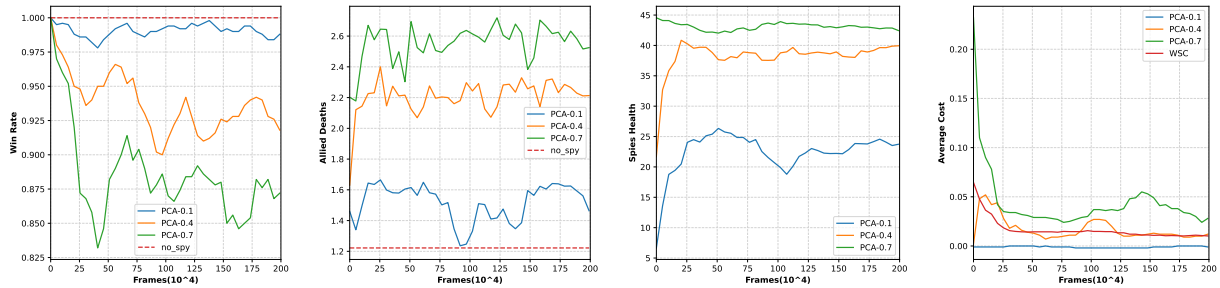
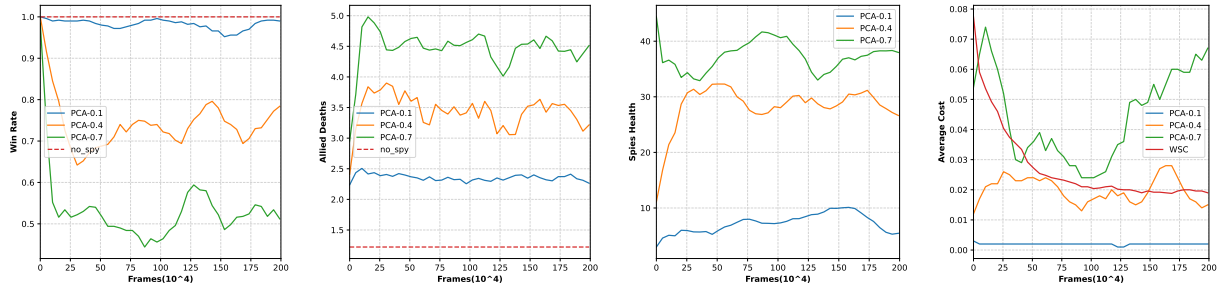


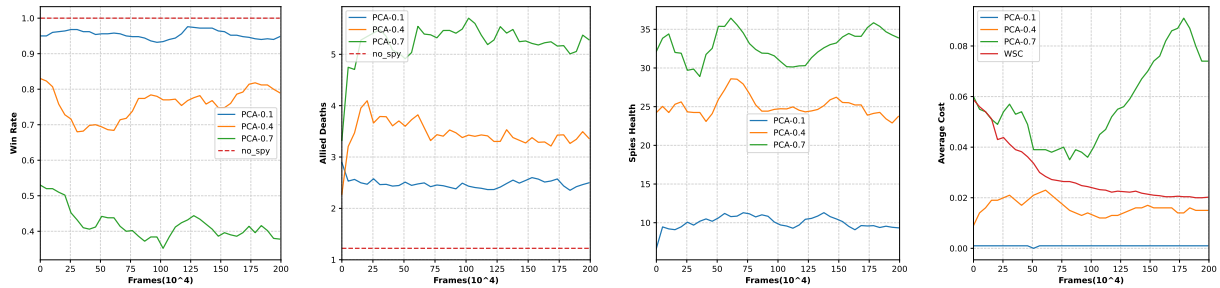
(a) Win rate. (b) Allied death. (c) Health. (d) Cost.

Figure 1: Adding one spy to the 6m-vs-6m environment in SMAC. (a) The allied win rate. (b) The allied deaths. To be noted, it includes the deaths of spies. (c) The average health of spies. (d) The average cost of spies.



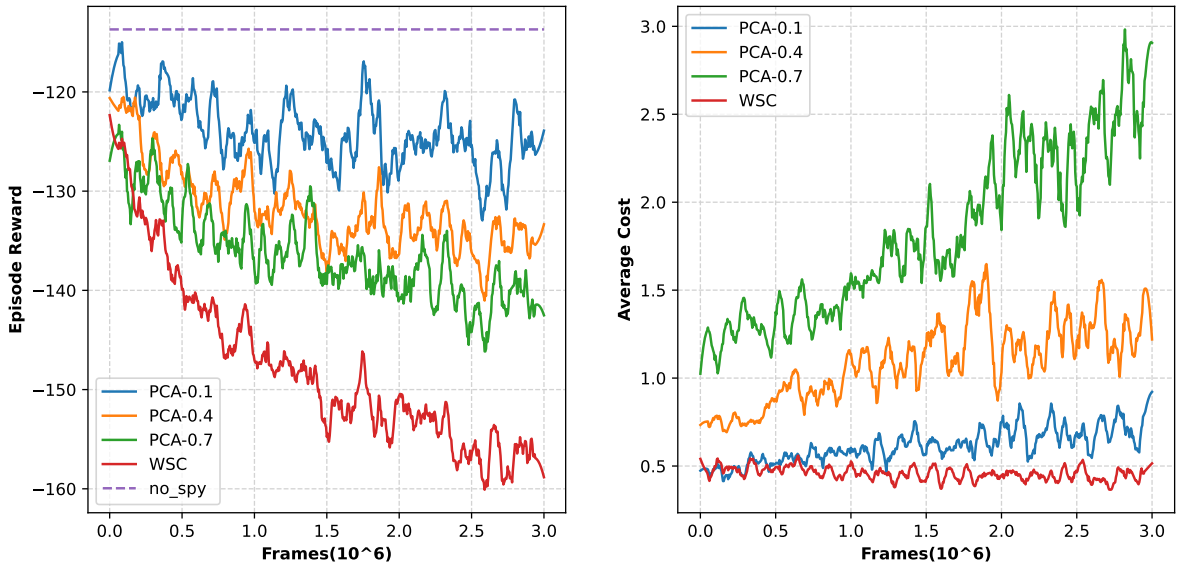
(a) Win rate. (b) Allied death. (c) Health. (d) Cost.

Figure 2: Adding two spies to the 6m-vs-6m environment in SMAC. The definition of subfigures corresponds to the subfigures in Fig.1, but with a different number of spies.



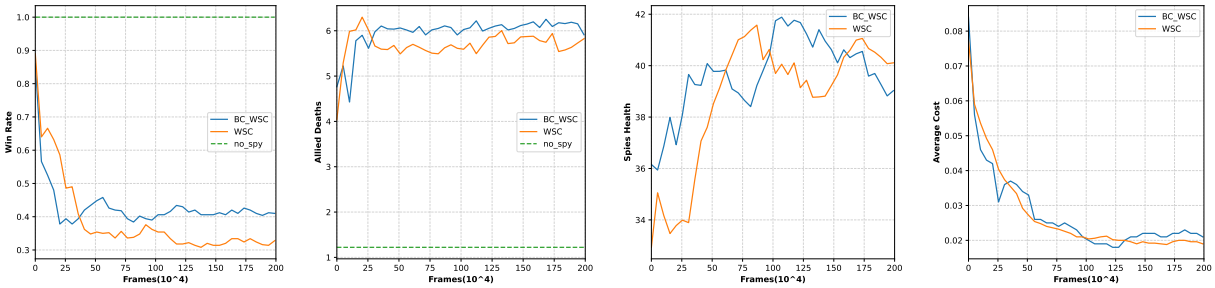
(a) Win rate. (b) Allied death. (c) Health. (d) Cost.

Figure 3: Adding three spies to the 6m-vs-6m environment in SMAC. The definition of subfigures corresponds to the subfigures in Fig.1, but with a different number of spies.



(a) Episode reward. (b) Cost.

Figure 4: Adding one spy to the simple_spread environment in MPE. The simple_spread environment has 3 agents, 3 landmarks. At a high level, agents must learn to cover all the landmarks while avoiding collisions. In this task, well-trained spies can reduce the rewards of victim agents by colliding and blocking.



(a) Win rate. (b) Allied death. (c) Health. (d) Cost.

Figure 5: We collect state-action pairs from the victim agents training process and use behavior cloning to obtain a cloned policy. This policy is then used to train the spy agents. By comparing this approach with the WSC method, we find that the cloned policy does not compromise the effectiveness or stealthiness of the attack.