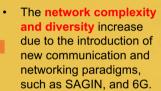


- In complex NGN scenarios, learning from expert trajectories can maximize the efficiency of policy refinement.
- The reward learned from IRL can maximize the effectiveness for mimicking expect trajectories (E).

Motivations for applying IRL



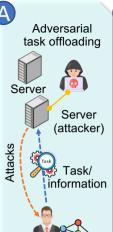
Environmental

complexity:

Numerous physical factors explicitly or implicitly affect environment, causing difficulty in reward modeling (D).

Reward unavailability:

- In many cases of NGN, the reward cannot be modeled, such as
- The network contains adversarial, e.g., attackers, eavesdroppers, whose objective is hidden from users (A).
- The immediate reward to each action is unavailable (B).
- The NGN is human-in-the-loop (C).



attacks

hidden objective

and strategies of

IRL infer the

attackers

IRL infer rewards containing subjectivity, thus aligning with human

