CAREER: Intermittent Learning Framework for Smart and Efficient Cyber-Physical Autonomy

Prof. Kyriakos G. Vamvoudakis

The Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology

Web: http://kyriakos.ae.gatech.edu/

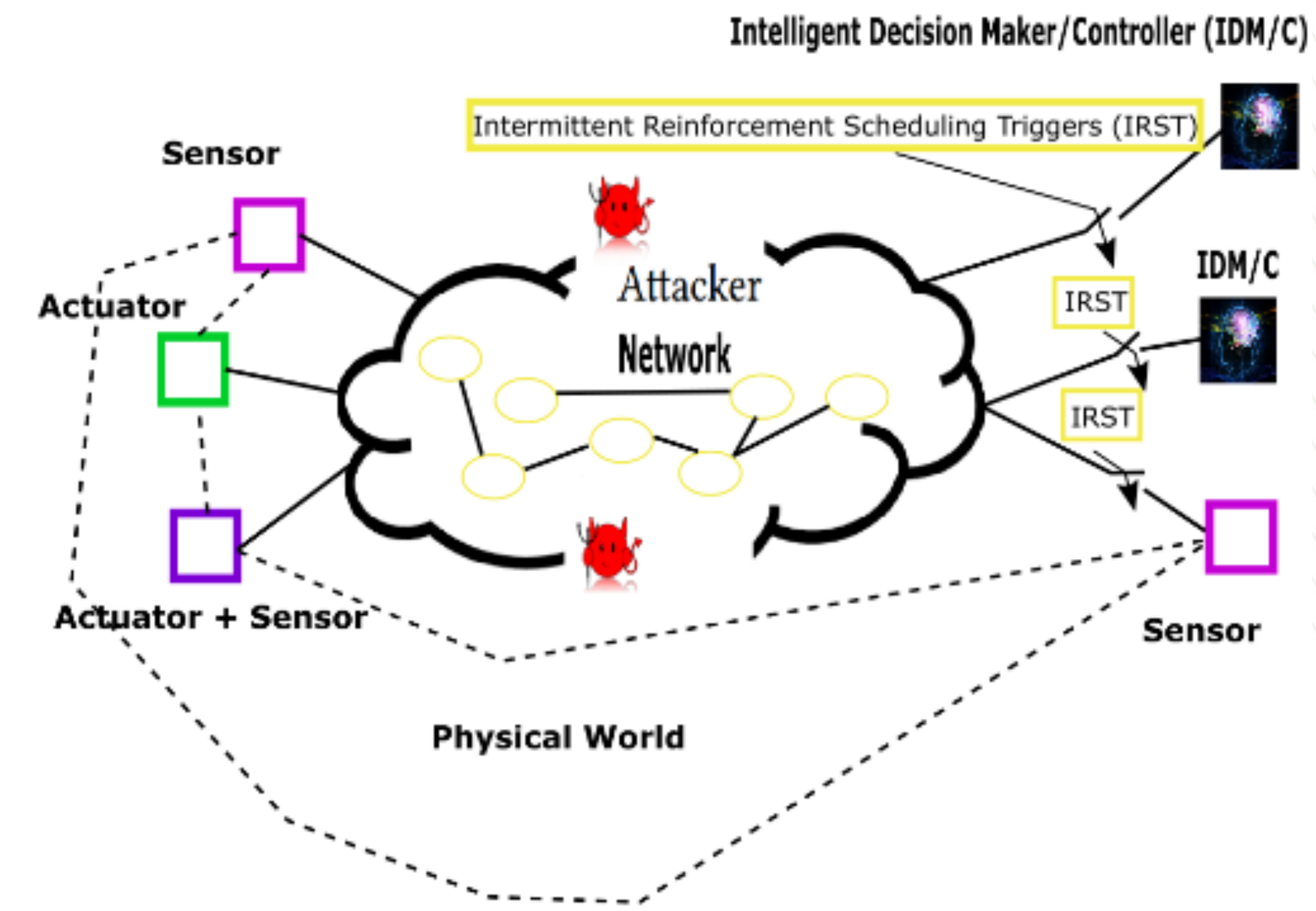
Project's Web: http://kyriakos.ae.gatech.edu/NSFCare

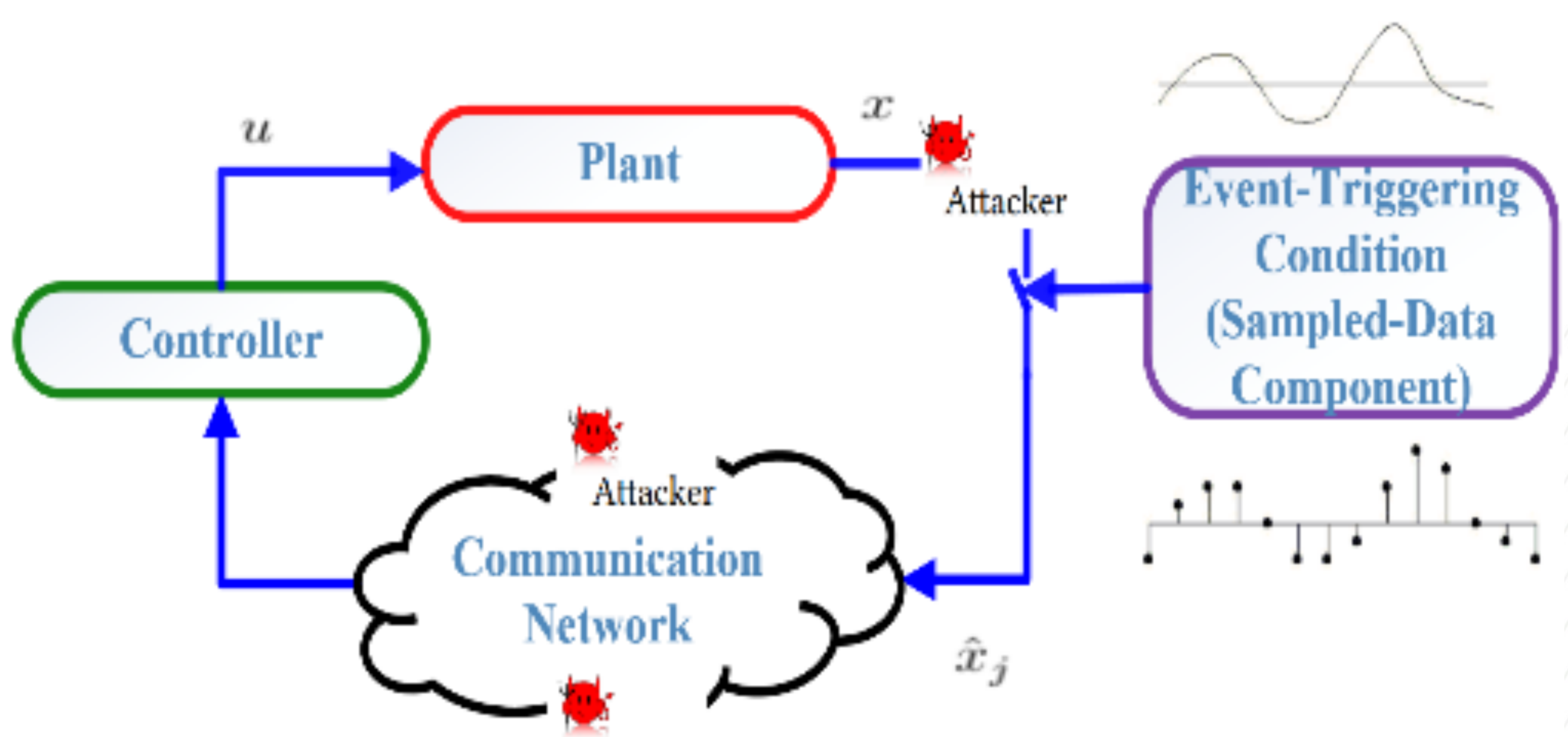
Email: Kyriakos@gatech.edu

ID: CPS-1851588

Intermittent Reinforcement Learning

- Drawbacks of existing learning mechanisms:
  - ➢Continuous data sharing
  - ➢Over utilization of communication channels
  - ➢Vulnerability to attacks
- Intermittent Data Sharing Schemes
  - Central nervous system
  - Q-learning + Intermittency
  - Error based triggers in update
- Key Impact Points:
  - Can we inform autonomous agents to operate in human-centric environments and increase their learning capabilities? Can we be model agnostic?
  - Can we mimic learning in humas? Can we generalize transferrable learning mechanisms between different components of CPS? Can we learn from other agents, humans or otherwise, by observation?
  - Can we leverage safe policies that are deployed intermittently to enforce secure learning in the presence of adversaries?



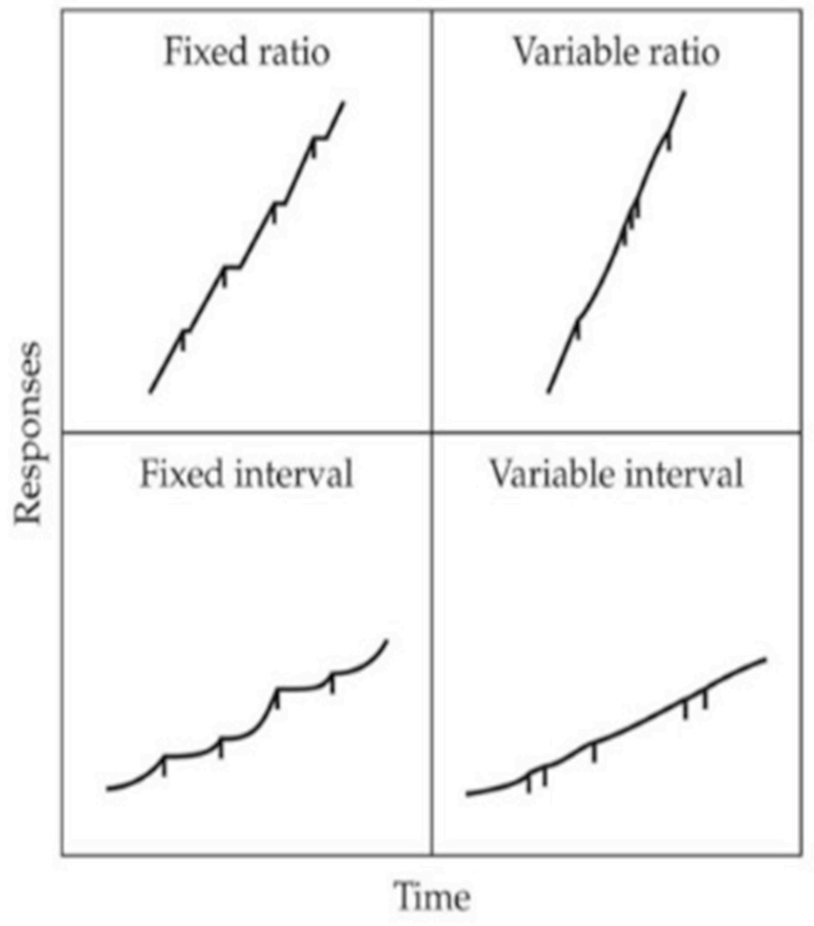


Intermittent Reinforcement Learning

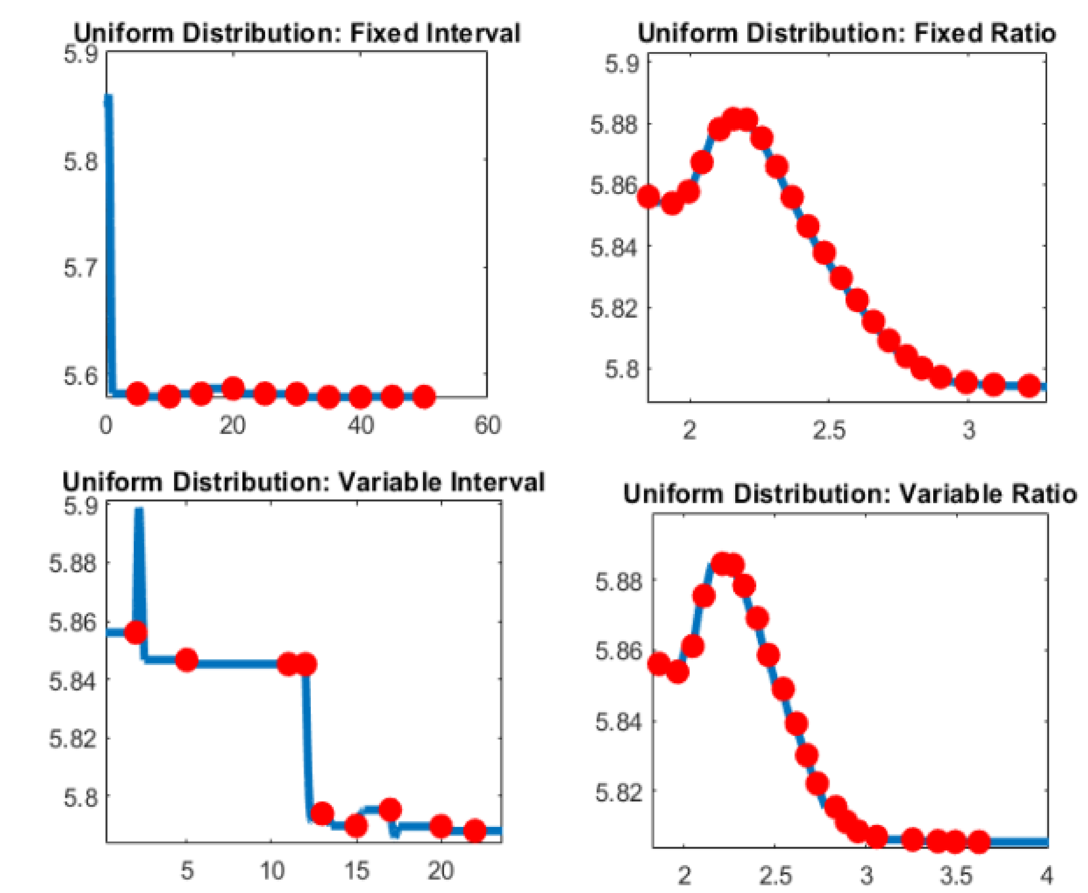
- Critic intermittency

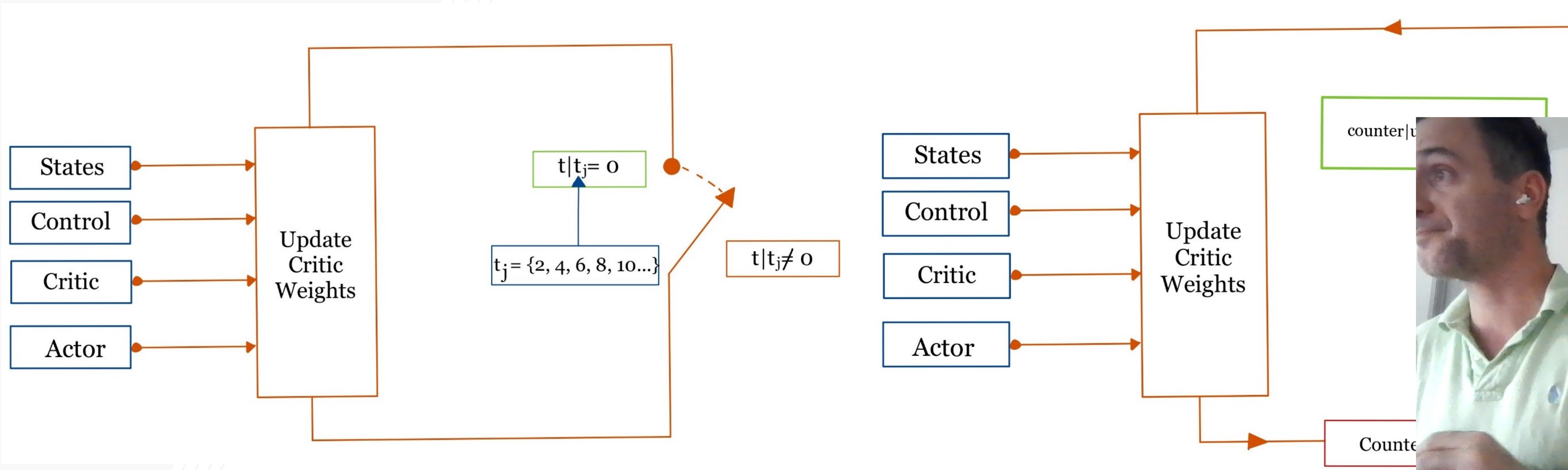
$$\dot{\hat{\theta}}_c(t) = -\alpha_c \frac{\sigma}{(1+\sigma^{\mathrm{T}}\sigma)^2} e_c, \rightarrow \begin{cases} \dot{\hat{\theta}}_c = ?, & t \neq t_j, \\ \hat{\theta}_c^+ = ?, & t = t_j, \end{cases}$$

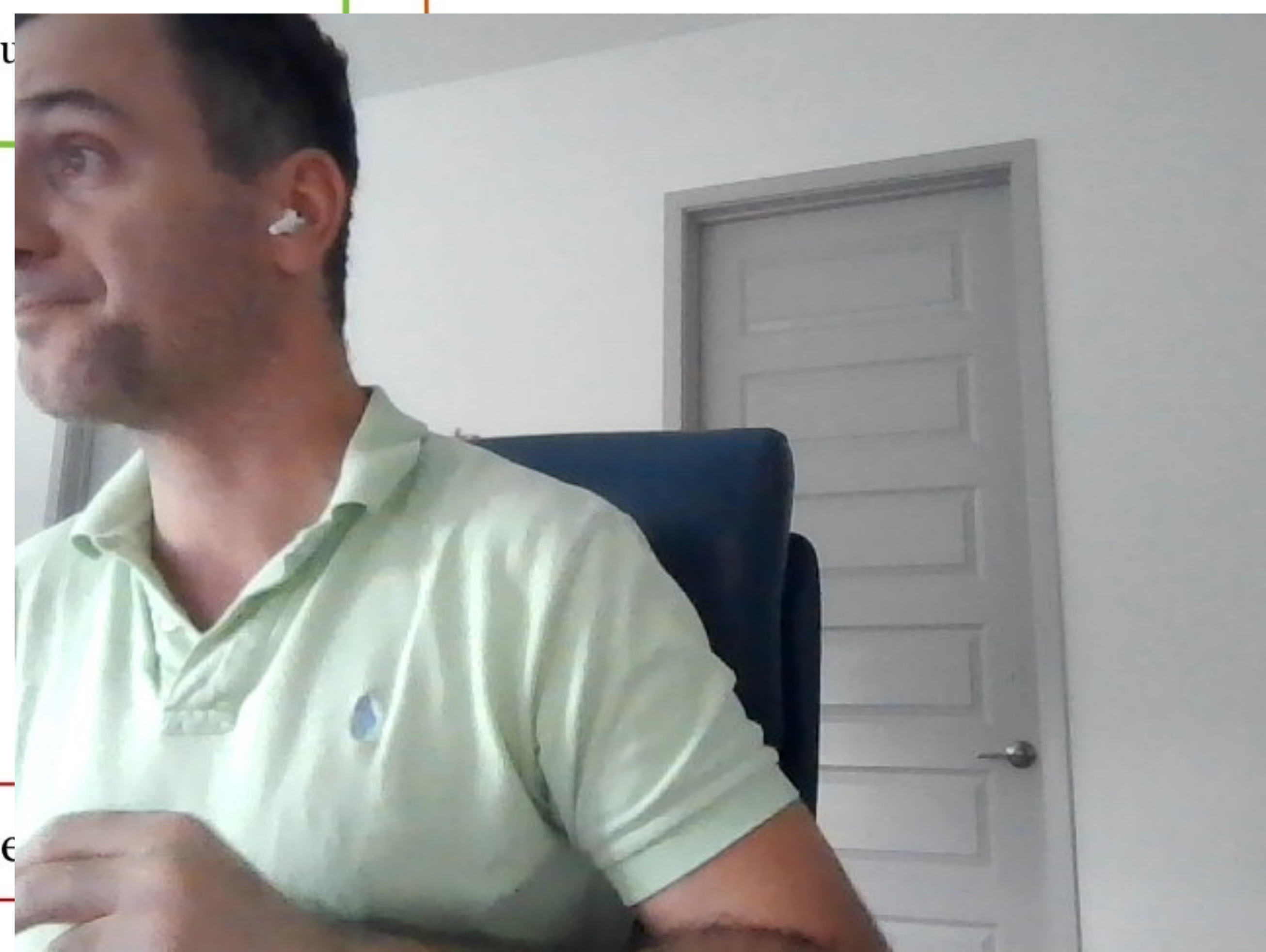
- B. F. Skinner’s operant conditioning ideation
- Schedules of intermittent rewards:
  - ➢Fixed interval schedule
  - ➢Variable interval schedule
  - ➢Fixed ratio schedule
  - ➢Variable ratio schedule



(b) Skinner’s response for variable/fixed reinforcements







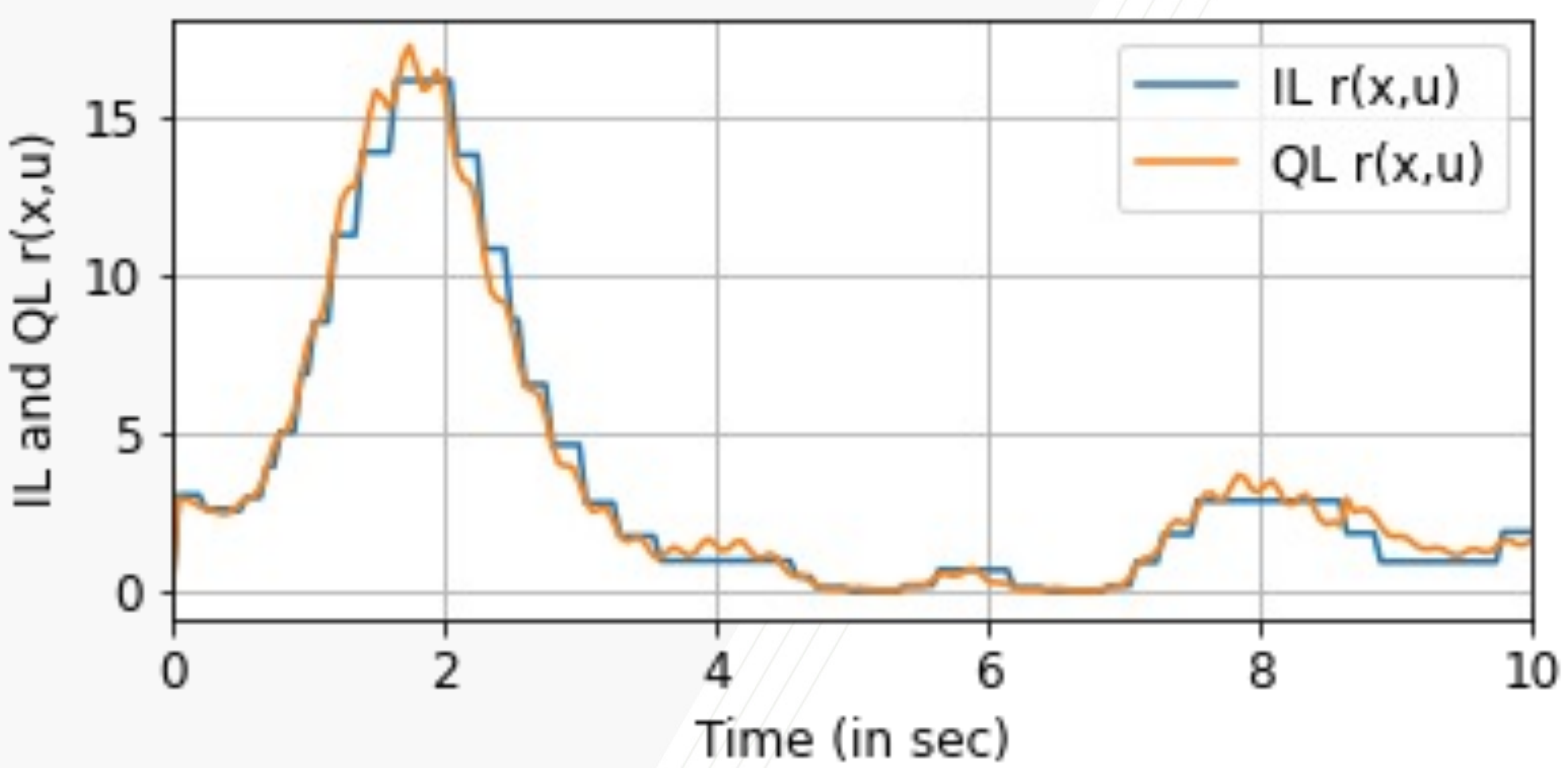
Intermittent Learning with Sparse Rewards

Sensing system breakdown, under the effect of reward loss

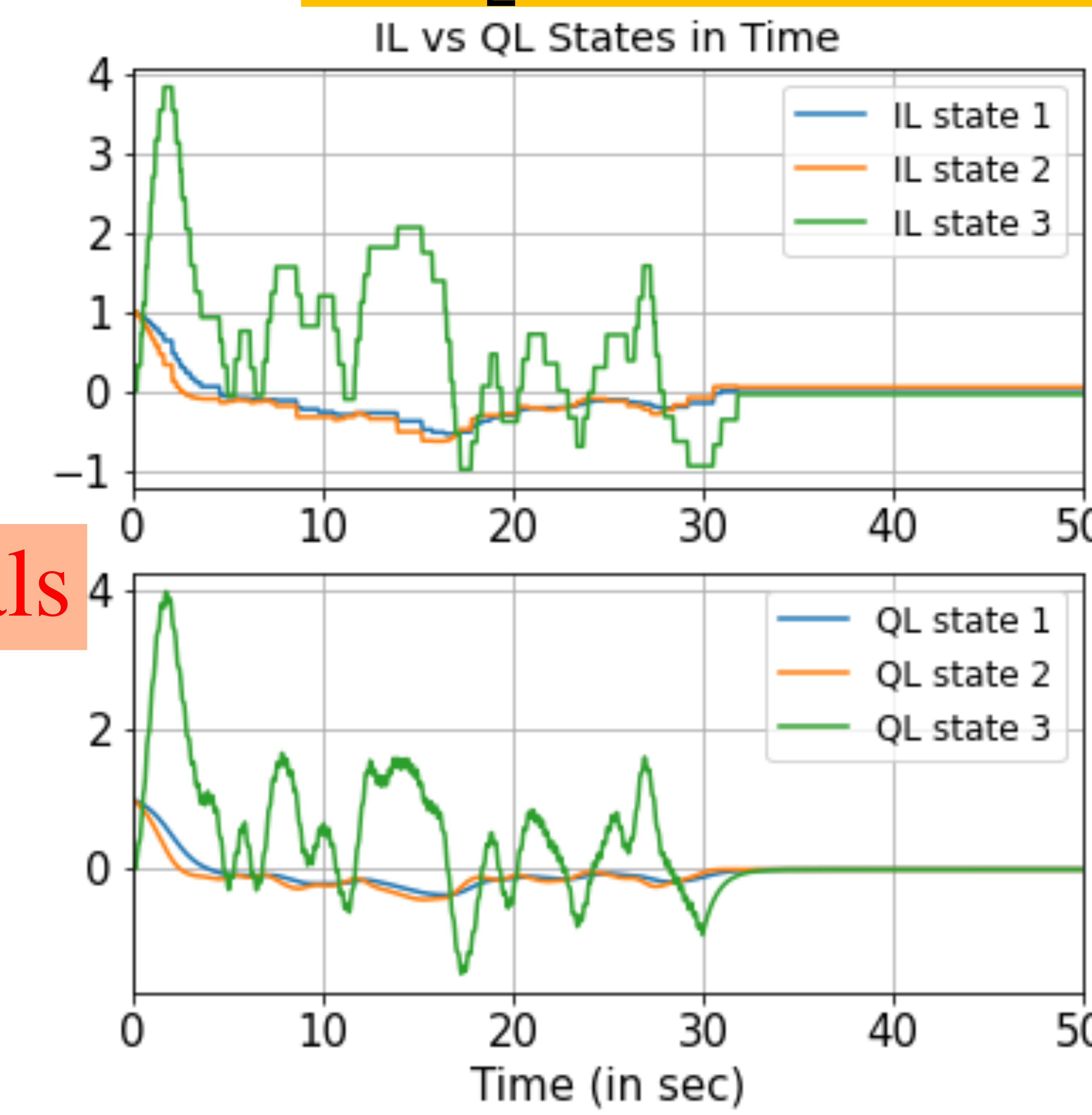
❖packet drops
❖jamming attacks

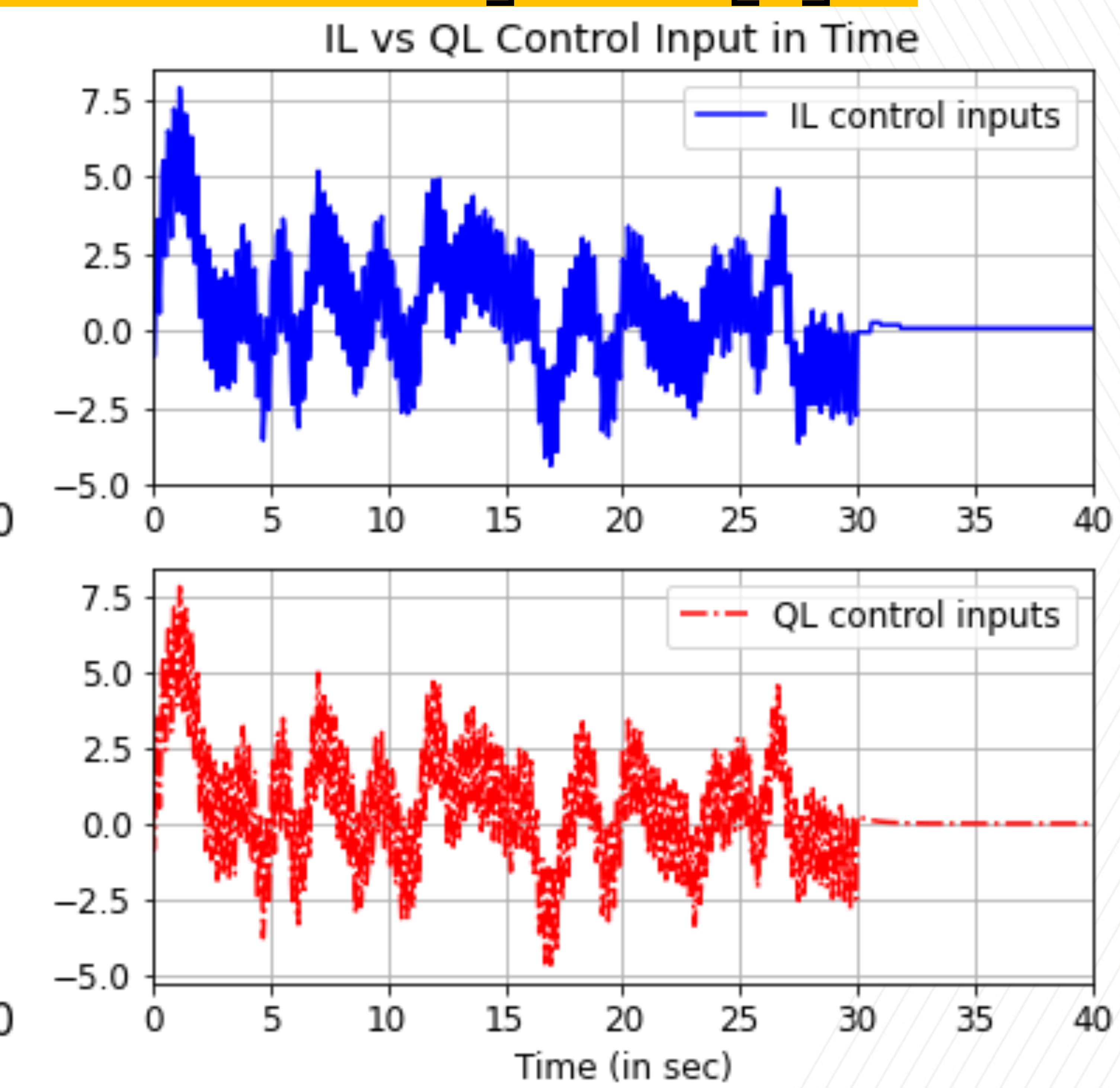
The lack of complete reinforcement signals

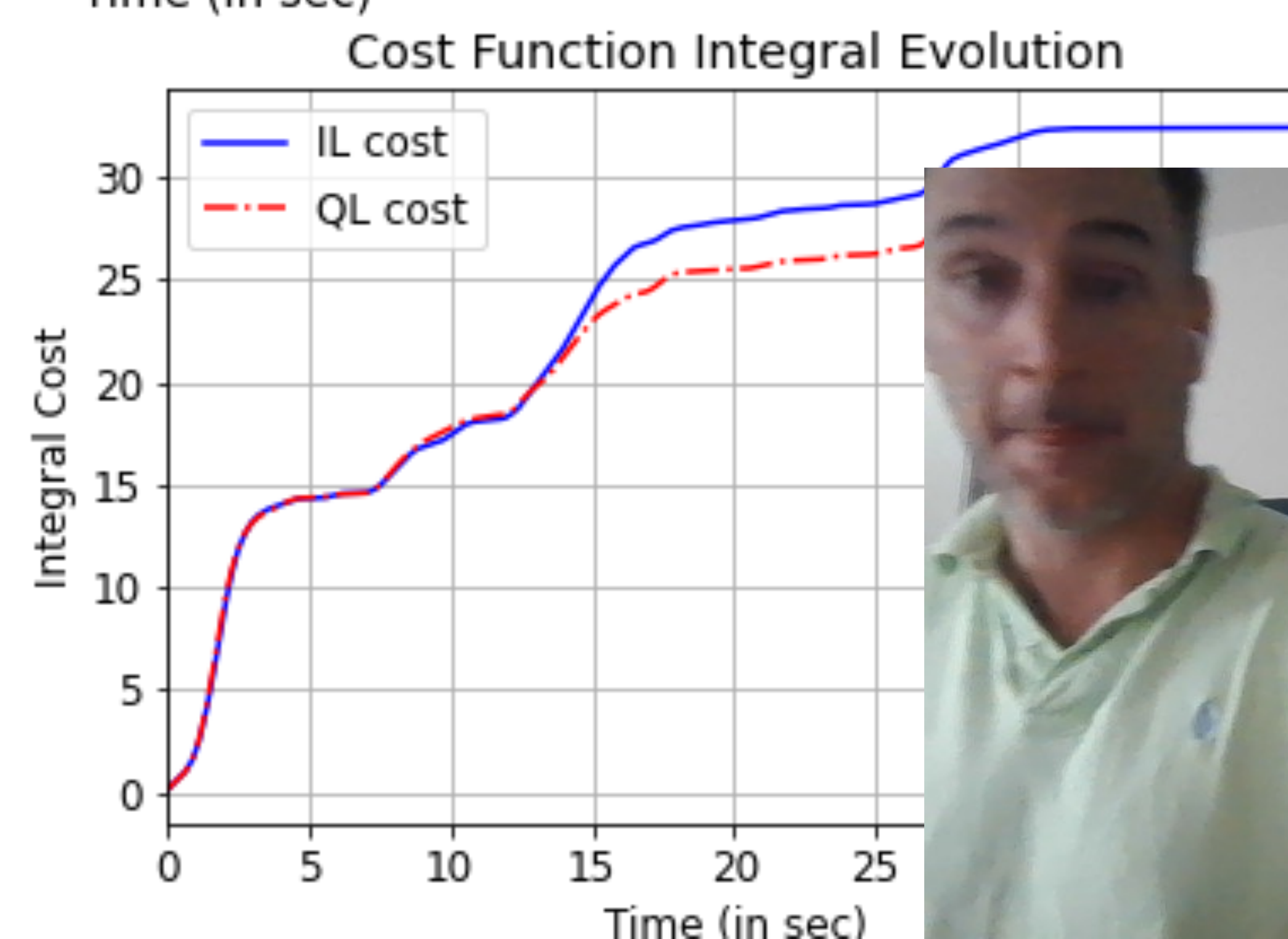
$$\dot{x} = \begin{bmatrix} -1.0189 & -0.9051 & -0.0022 \\ 0.8223 & -1.0774 & -0.1756 \\ 0 & 0 & -1.0000 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u$$

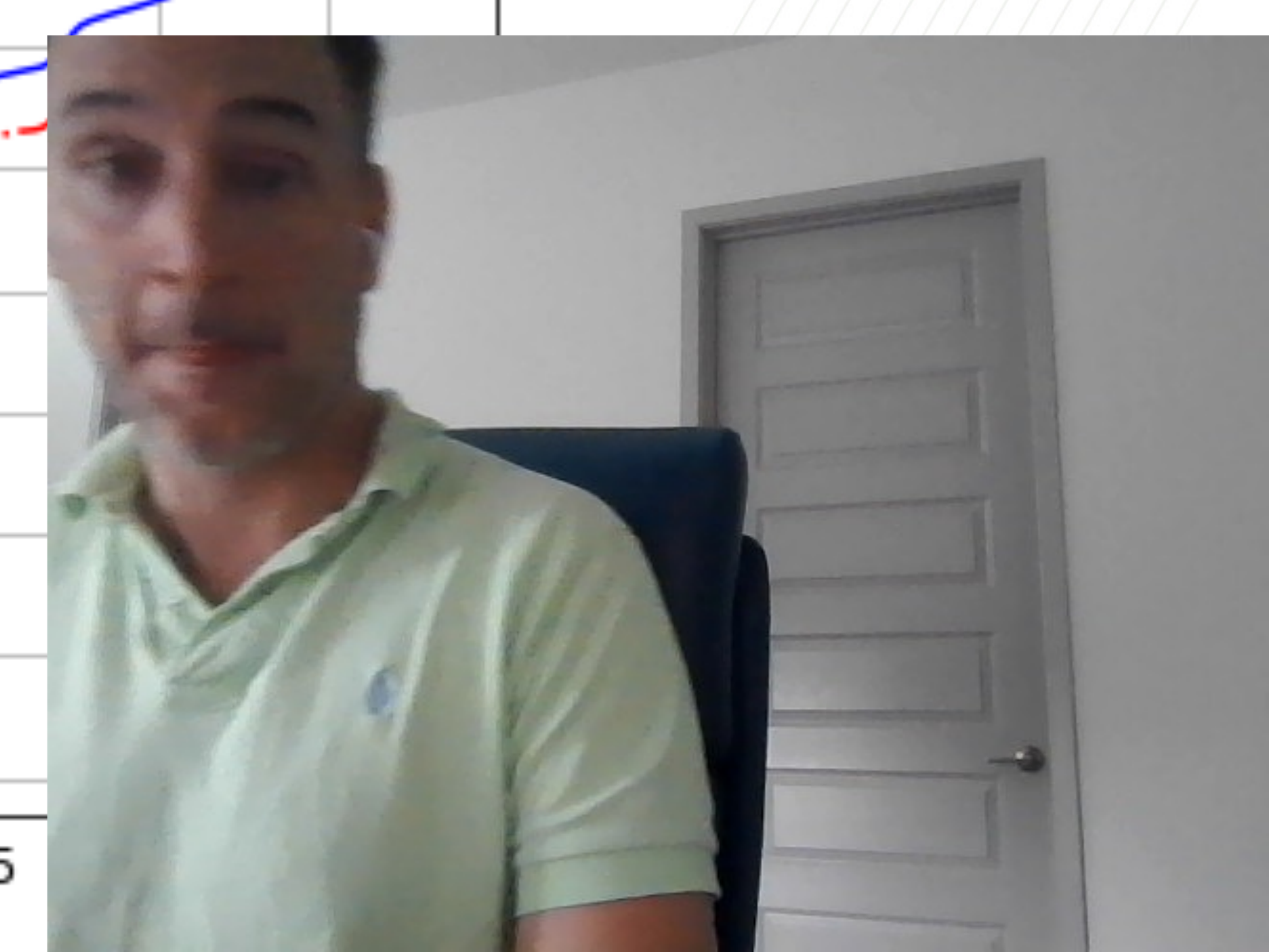


$$p(t) = \begin{cases} 1, & r(x,u) \text{ is transmitted} \\ 0, & \text{otherwise.} \end{cases}$$





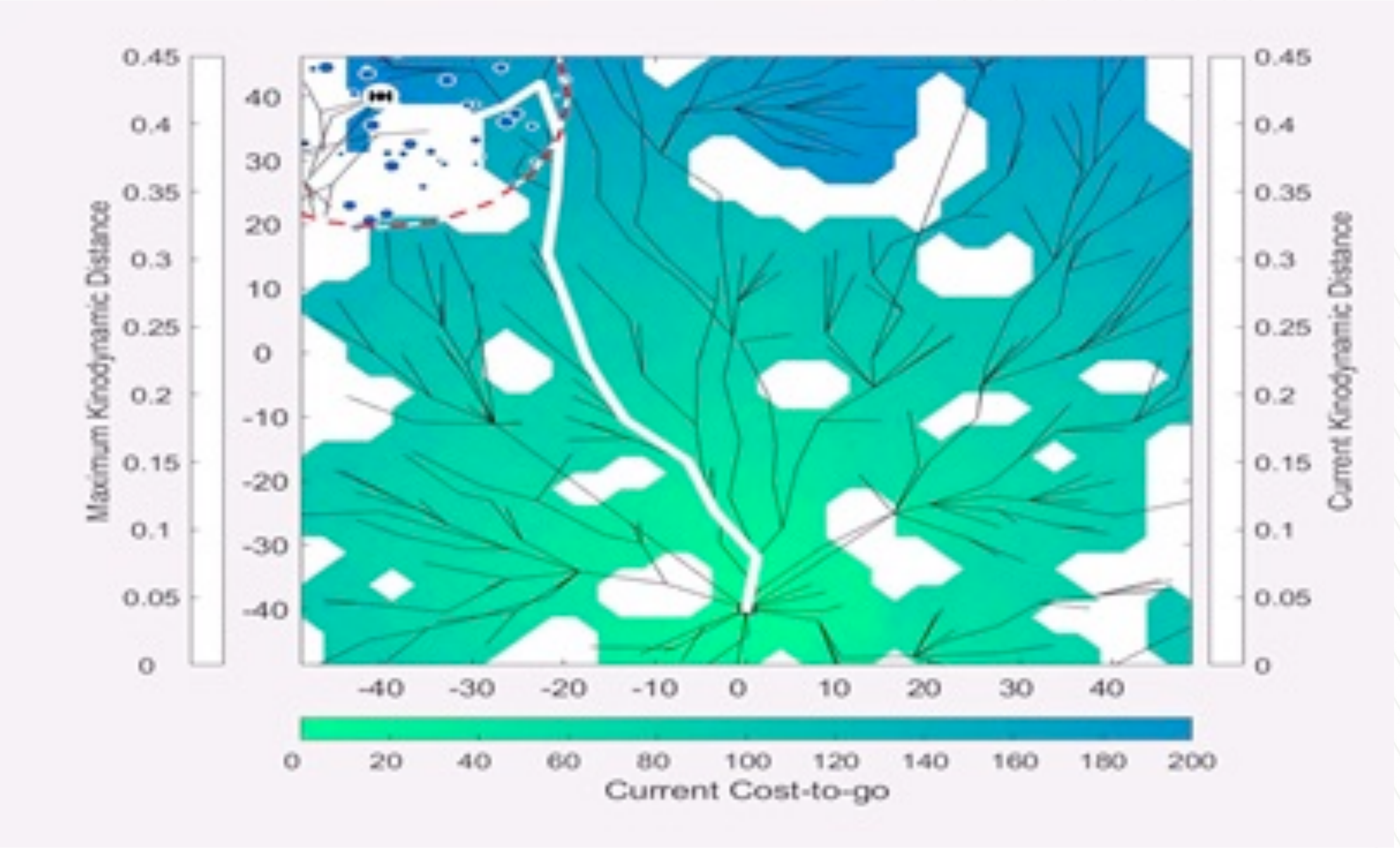


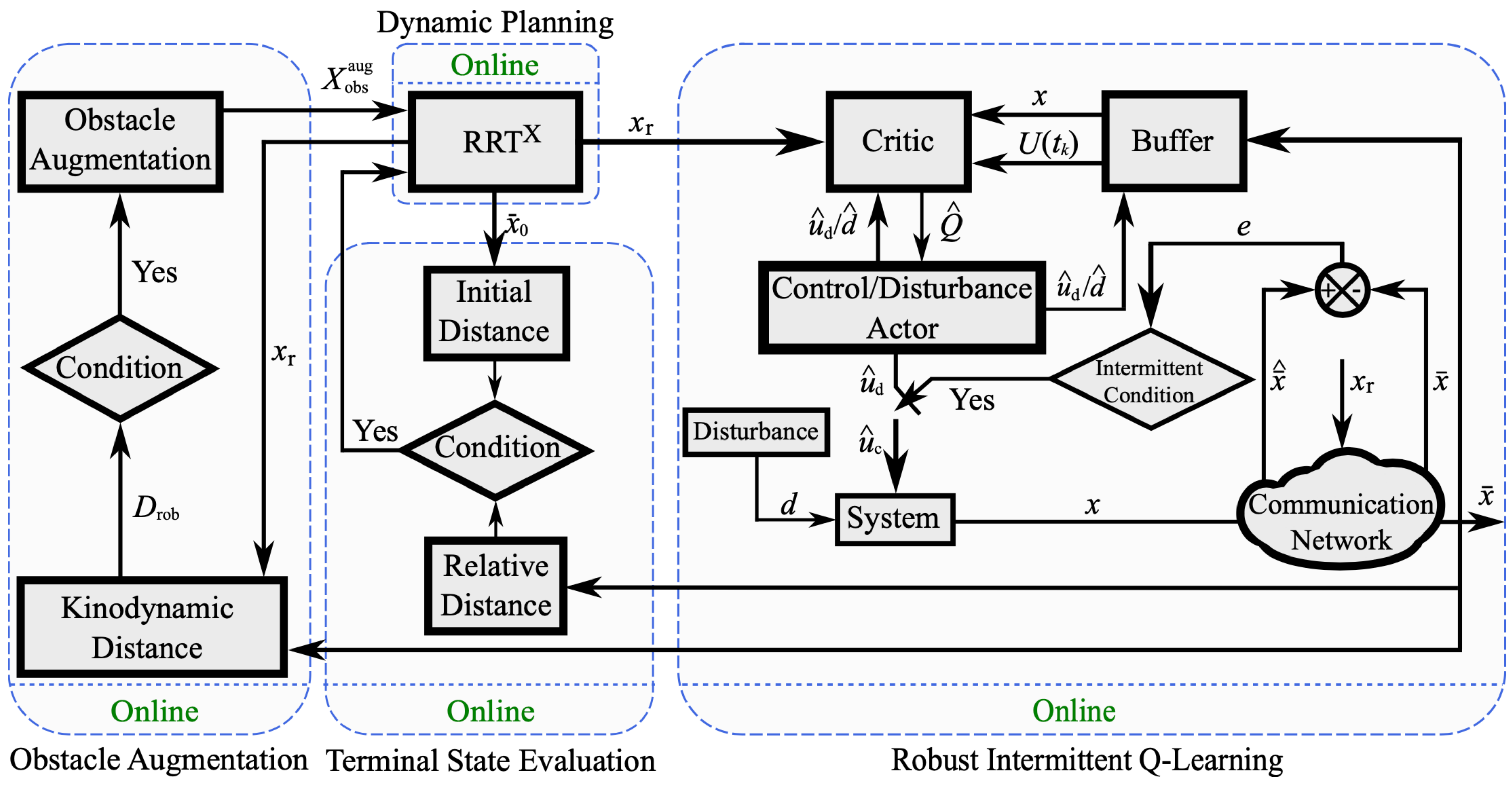


Intermittent Model-Free Learning-based Motion Planning

- Intermittent transmission of control inputs to minimize communication overhead

- A relaxed exploration technique is employed to improve the convergence speed of the intermittent Q-learning





By limiting the squared norm error with a calculated threshold, the intermittent control policy guarantees stability of the equilibrium point and
computati
communic