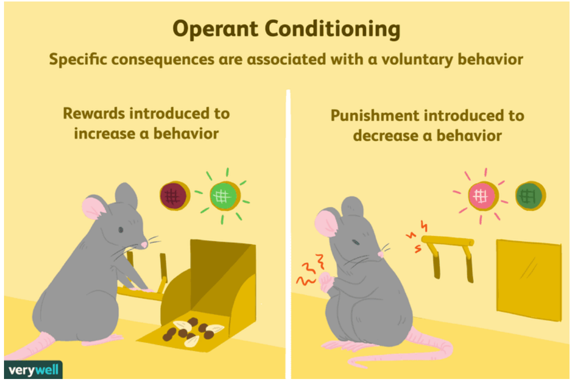
CAREER: Intermittent Learning Framework for Smart and Efficient Cyber-Physical Autonomy

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/NSFCareer.html
Email: Kyriakos@gatech.edu
ID: CPS-1750789, CPS-1851588

Why Intermittent Learning in CPS?

Allow full CPS autonomous operation in the face of unknown, bandwidth restricted, and adversarial environments

- The schedules of intermittent learning are either based on time (interval) or on behaviors (ratio) and can be fixed or variable.
  - Fixed-interval schedule is when a behavior is rewarded after a set amount of time.
  - Variable-interval schedule, is when a CPS agent gets the reinforcement based on varying and unpredictable amounts of time.
  - Fixed-ratio schedule, is when there are a set number of responses that must occur before the behavior is rewarded.
  - Variable-ratio schedule, is when the number of responses needed for a reward varies.

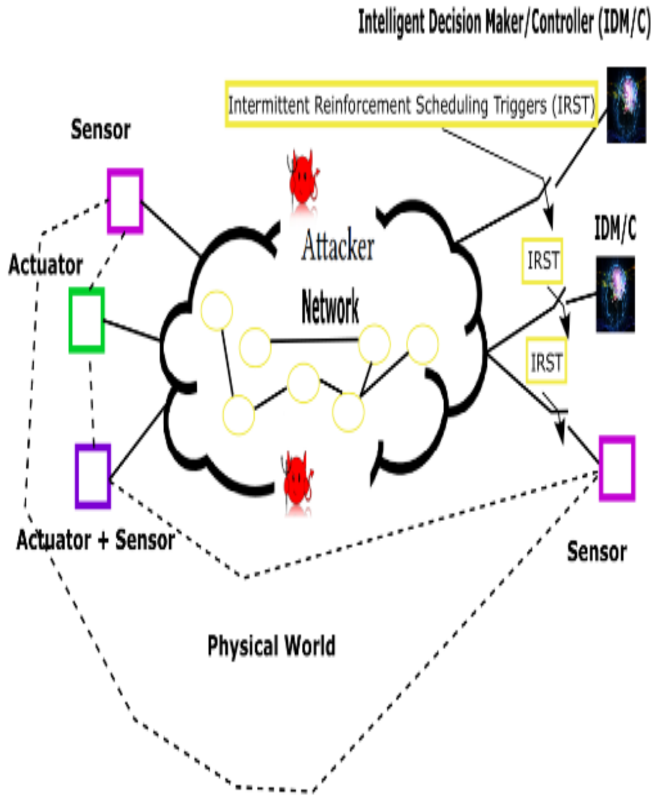


This distinction in the quality of performance can help determine which reinforcement method is most appropriate for a particular CPS situation; fixed ratios are better suited to optimize the quantity of output, whereas a fixed-interval can lead to a higher quality of output.

Description

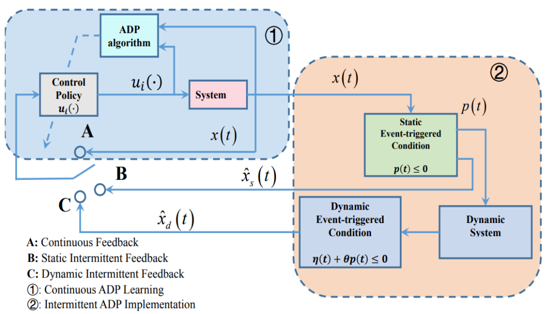
Goals of This Project:

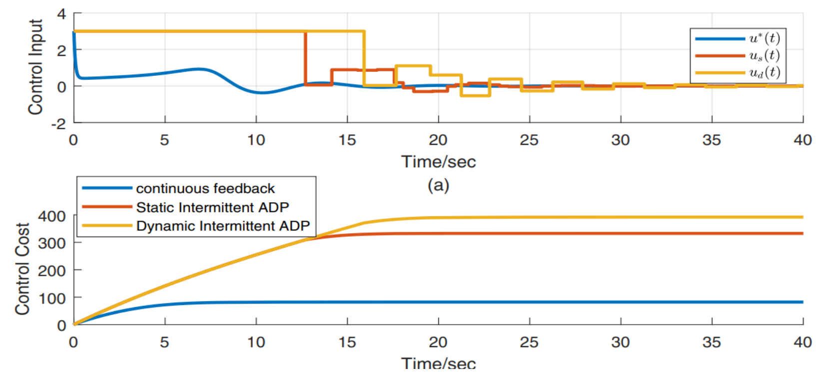
- How can we incorporate and fully adapt to totally unknown, dynamic, and uncertain environments with intermittent learning?
- How do we co-design the action and the intermittent schemes? How can we provide quantifiable real-time performance, stability and robustness guarantees by design?
- How do we solve congestion and guarantee security?

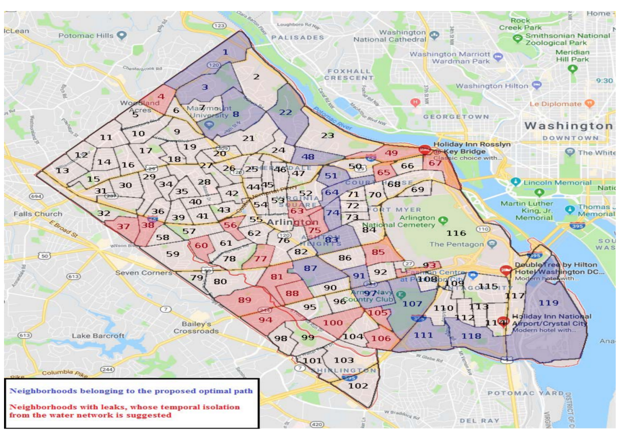


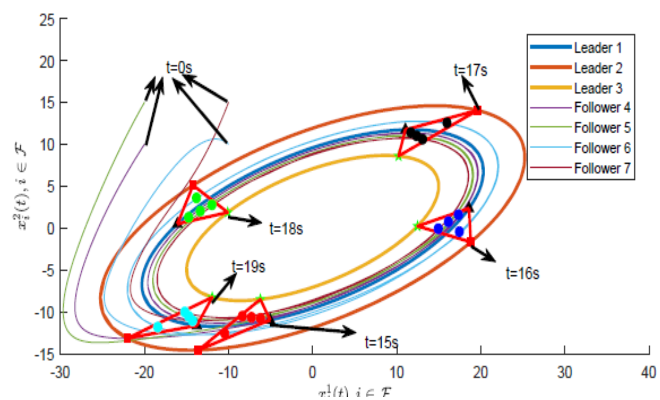
Findings

- Static and intermittent learning through approximate dynamic programming
- Dynamic intermittent feedback design For containment control
- Predictive intermittent Q-learning with application to CPS/IoT



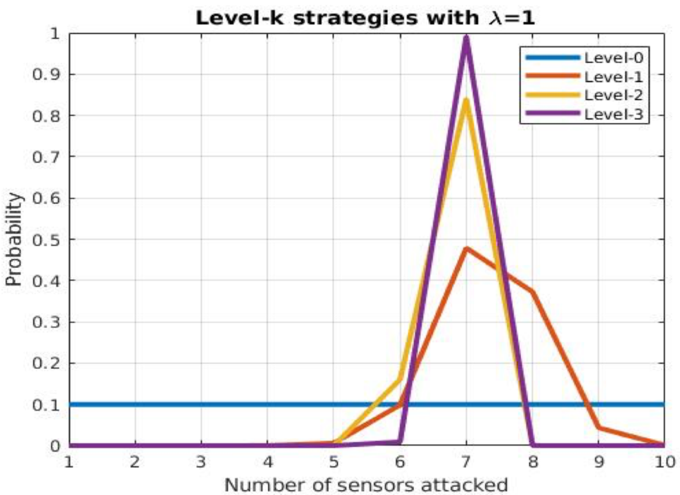






Findings

- Learning the levels of intelligence
- The limiting conditions as the cognitive levels increase, as well as when the CPS agents fully coordinate, are shown to converge to the Nash equilibrium.
- Each level-k agent behaves based on intermittent and subjective beliefs of the others' behaviors.



Algorithm 1: Intelligence Level Learning

1: **procedure**
2: Given initial state x_0, cost weights M, R, γ, highest allowable level defined to be $\mathcal{K}$ and time window T_{IRL}.
3: **for** $k = 0, \ldots, \mathcal{K}$ **do**
4: Set $j := u$ to learn the level-k defender policy.
5: Start with an initial guesses for $\hat{W}_u^k$, $\hat{W}_{u,a}^k$.
6: Propagate the augmented system with states $\chi = [x^{\mathrm{T}} \quad \hat{W}_u^{k\mathrm{T}} \quad \hat{W}_{u,a}^{k\mathrm{T}}]^{\mathrm{T}}$
7: Set $j := d$ to learn the level-k adversarial policy.
8: Start with initial guesses for $\hat{W}_d^k$, $\hat{W}_{d,a}^k$.
9: Propagate the augmented system with states $\chi = [x^{\mathrm{T}} \quad \hat{W}_d^{k\mathrm{T}} \quad \hat{W}_{d,a}^{k\mathrm{T}}]^{\mathrm{T}}$ Go to 3.
10: **end for**
11: Define the interaction time with each adversary as T_{int}, the number of total interactions n_{int} and an initial guess for λ.
12: **for** $i = 1, \ldots, n_{\mathrm{int}}$ **do**
13: For $t \in [t_i - T_{\mathrm{int}}, t_i]$, measure the value
14: Compute the mean level
15: Update λ Go to 13 to interact with a different adversary.
16: **end for**
17: **end procedure**

Findings

- Proactive defense - Probabilistic switching combining overall uncertainty/optimality for non equilibrium intermittent learning.
- Reactive defense – Isolate the suspicious learning components.