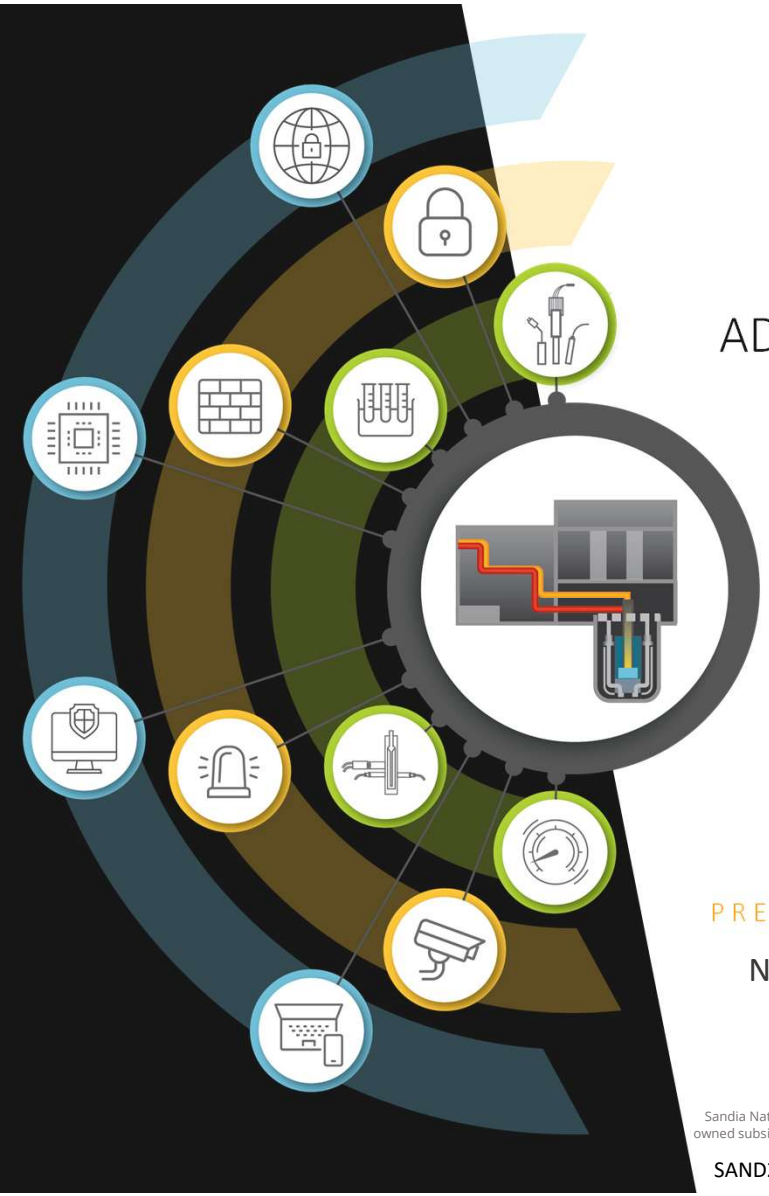


ADVANCED REACTOR SAFEGUARDS & SECURITY

Reinforcement Learning for Improved PPS

Spring FY26 Update

PRESENTED BY

Nathan Shoman

4/29/2026

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2026-20160PE

Motivation / Recap

Facility physical protection design can be slow and expensive

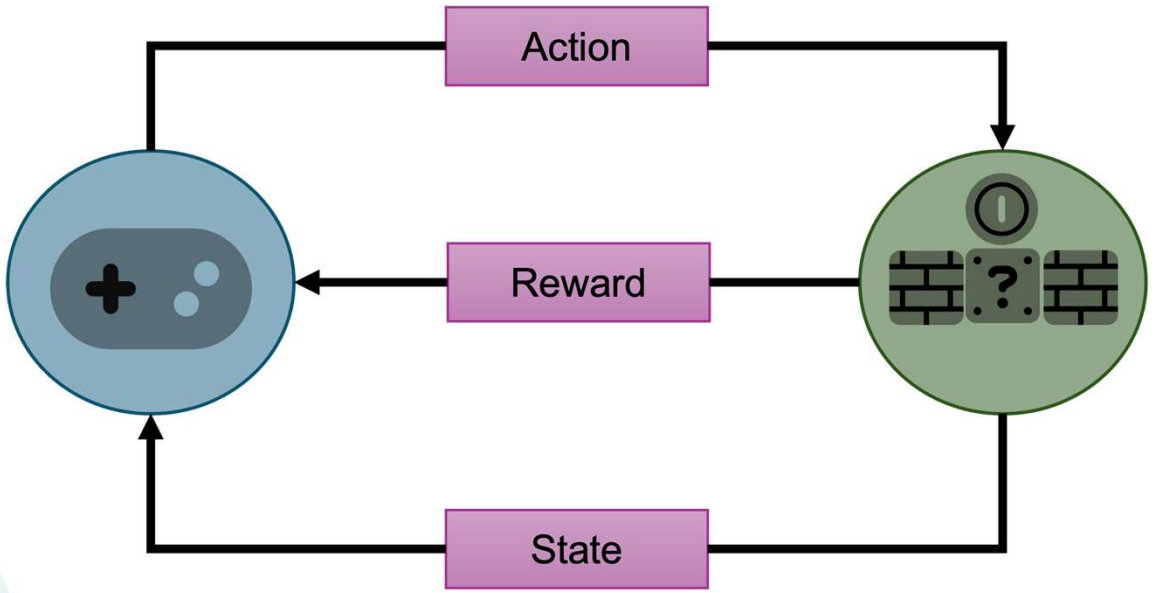
- Physical security can be a significant component of operation and maintenance costs for nuclear power plants
  - Consequently, optimizing for costs while retaining effective security is an ongoing development priority
- Designing physical protection systems (PPS) can take considerable time and rely on expert judgement
- PPS design can be thought of as a large-scale optimization problem
- New approaches and tools could accelerate the development cycle and resulting in cheaper, but more effective designs

Agent-based exploration can help more thoroughly explore the PPS design landscape

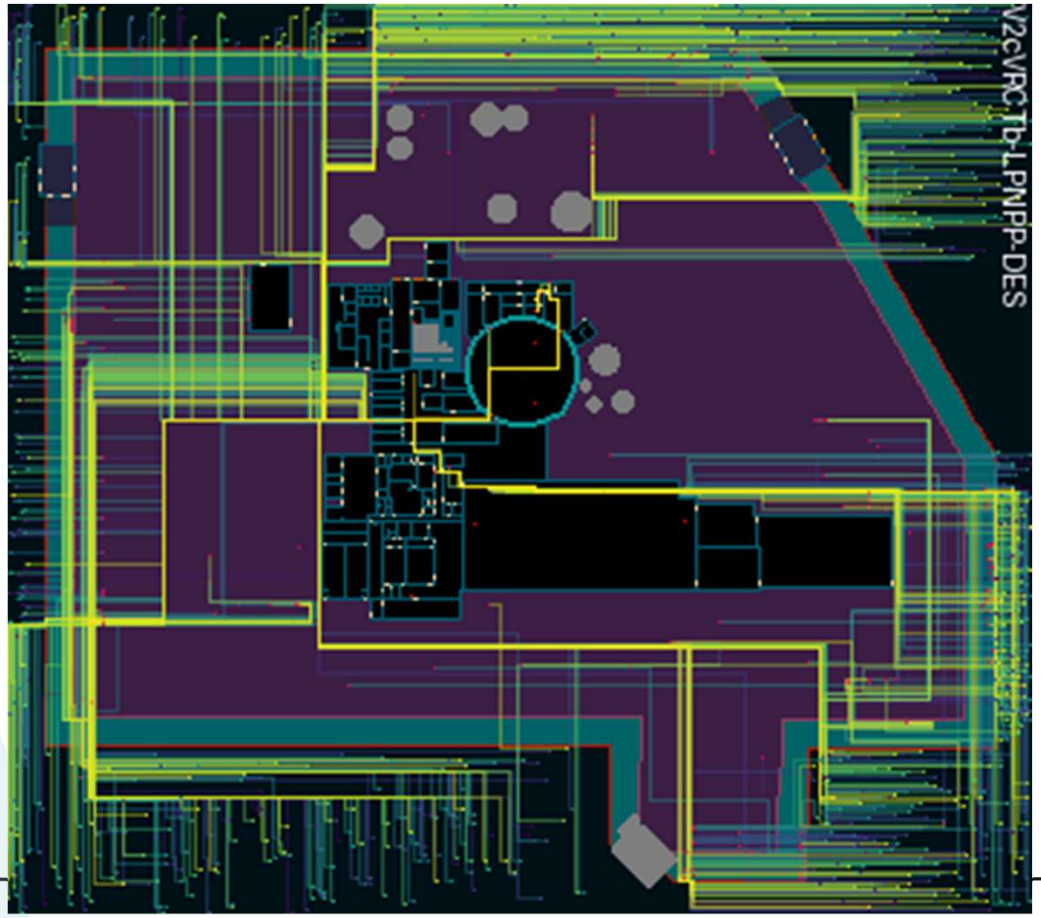
- **Benefit**: More thoroughly explore the design space
- PPS design is a complex problem relying on expert judgement combined with trial and error (i.e., system evaluations)
- Reinforcement learning can bring a principles exploration of the design space to the problem

Image credit: Losslandscape

Learn by playing: reinforcement learning

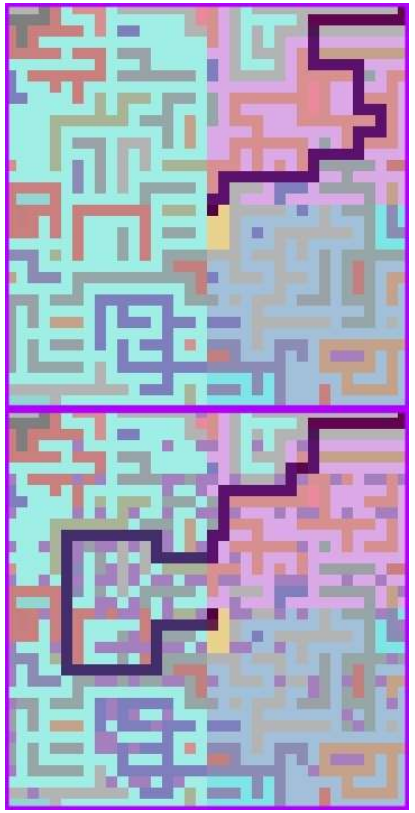


Previously: Agents learn the most vulnerable path entirely through self-play



Agents can learn design too, but perhaps less useful

- Agents can learn to use a mix of both detection and delay to increase path length
- Can surprisingly generalize to unseen designs
  - Perhaps there's some hidden commonality in random mazes? Unclear.
- Much denser reward than adversarial problem; quick to learn hard to master
- Flat agent outperform hierarchical agent, but unclear why

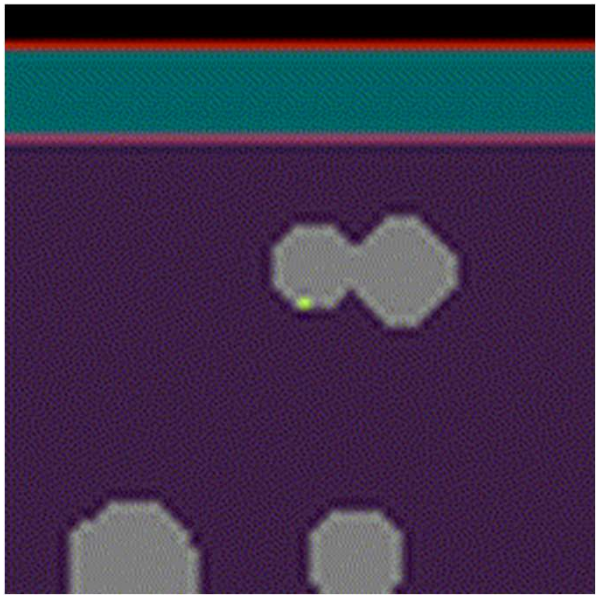


Focus on improving TRL of adversarial approach while adding new features

- Discussed with stakeholders the relevance of different approaches developed so far
- Design tasks less relevant/important than adversarial behavior discovery
- Efforts in FY26 have focused on implementing new behaviors and capabilities not see in current state-of-practice tools

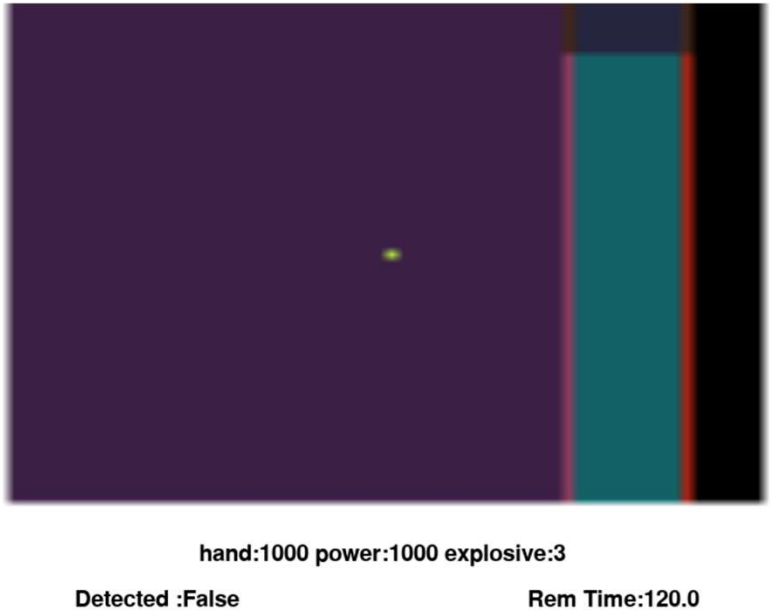
Environmental destruction and debris

- Agents can optionally destroy the environment, permanently altering the space.
- Certain tools can be specified to cause debris
  - Debris can modify either movement speed or vision



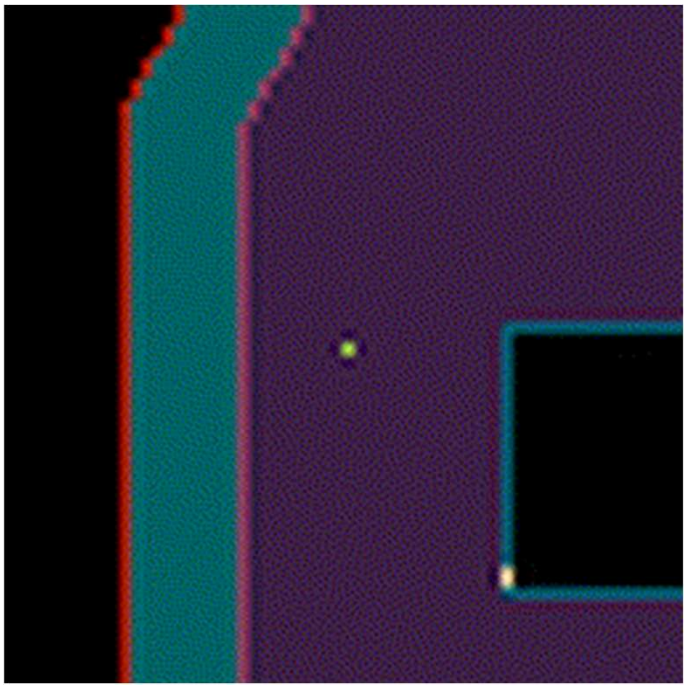
Limited tool use

- Agent's tool use can be limited by type
- Dynamic changes in probability of detection due to sensing element changes
  - Agent could reduce microwave cross section by using certain tools earlier
- Planning: Dynamic agent tool selection, tool "costs" and "loadouts"



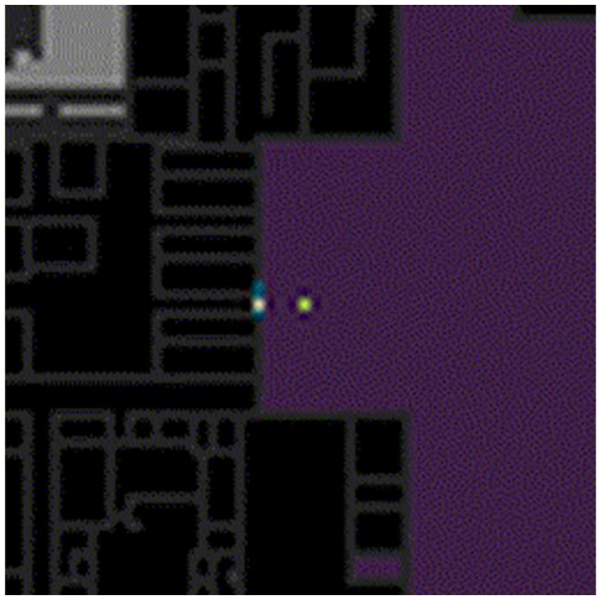
Target area fog

- Fog-of-war type obscuration can be applied to arbitrary areas
- Hides everything in the area, including sensing elements and targets
- Supports a partial knowledge adversary



General uncertainty

- Fog can be combined with barrier uncertainty
- Applicable to cases where facility layout is known, but PPS elements are not
- Can be applied to areas or entire map



Zero-shot performance

- Popular question: Can we see the impact of small changes or evaluate on unknown/unseen facilities?
- Not yet, this framing goes beyond RL and other complementary techniques are needed
- RL only learns state-action transitions from a known-ish environment and would need test-time exploration

$$V^{\pi*}(s) = \max_{a} \left\{ R(s,a) + \gamma \sum_{s'} P(s'|s,a) V^{\pi*}(s') \right\}$$

Bring in the robots

- Common task in robotics is exploring new environments
- Can borrow from existing techniques like SLAM or Hydra
- Challenge: Real-time exploration techniques usually framed for 3D environments, not 2D

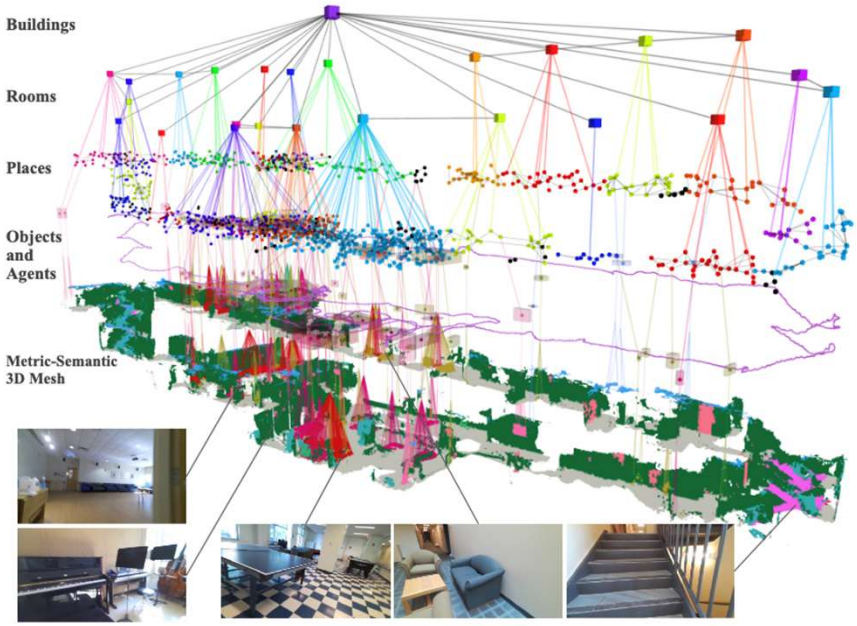


Image credit: Huges et al. 2022

Ongoing FY26 Work

- Developing interpretable performance metrics
  - Existing state-of-practice metrics like critical detection point don’t map as well to the RL adversary
- Journal paper development
  - Two journal papers for each approach (adversary, designer) being developed
- Generating results and insights from new features

Questions?