Sandia National Laboratories

Exceptional service in the national interest

DUAL-ROLE AGENTS FOR IMPROVED PHYSICAL SECURITY DESIGN

Nathan Shoman

NRC AI Workshop 2025

September 2025

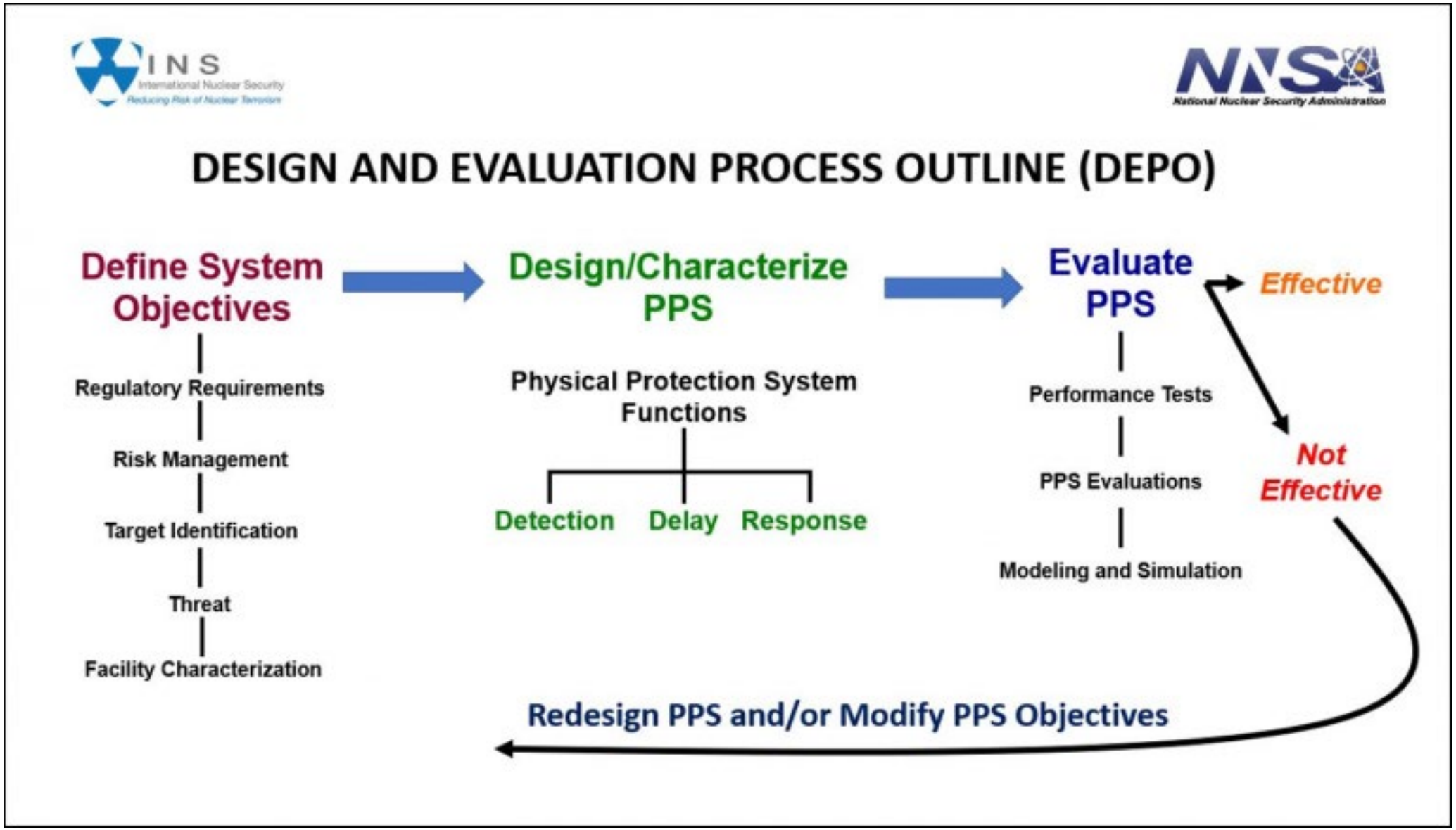
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2025-11641PE

FACILITY PHYSICAL PROTECTION SYSTEM DESIGN CAN BE SLOW AND EXPENSIVE, BUT IMPORTANT

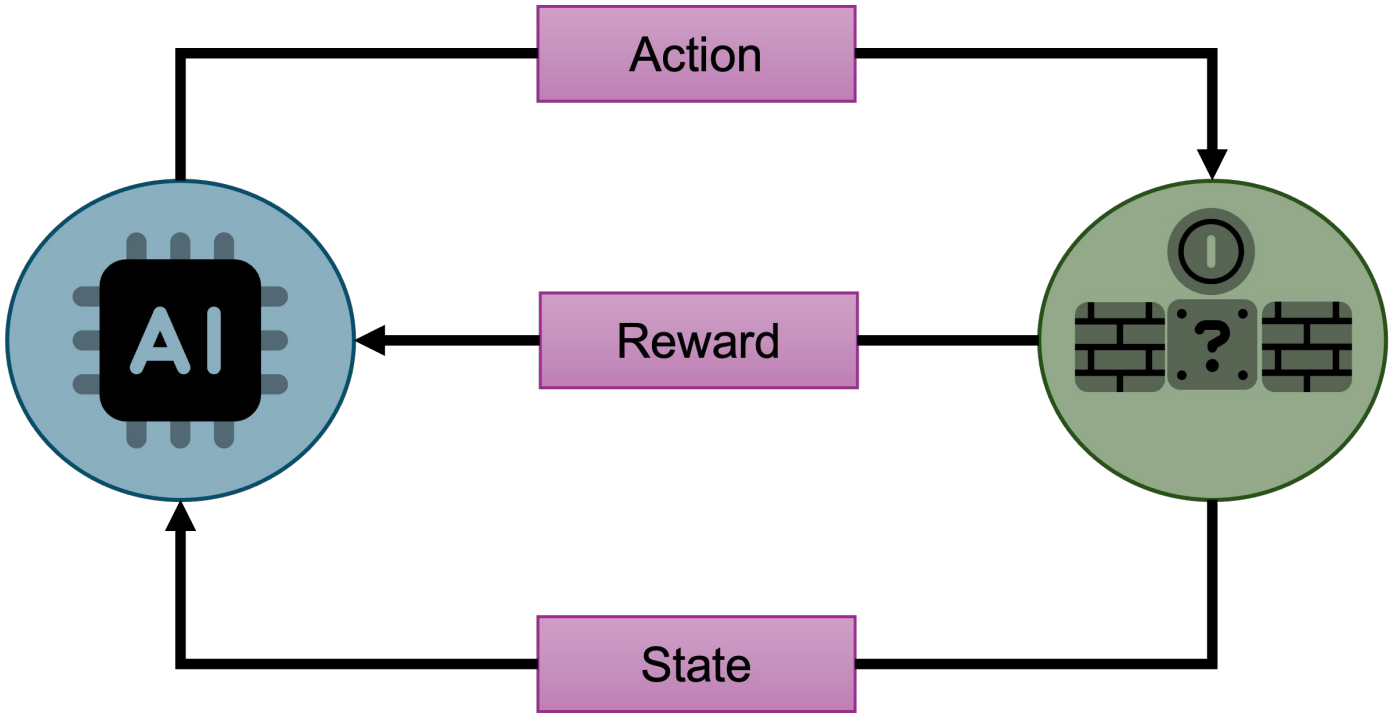
- Physical security can be a significant component of operation and maintenance costs for nuclear power plants
  - Consequently, optimizing for costs while retaining effective security is an ongoing development priority
- Designing physical protection systems (PPS) can take considerable time and rely 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

DESIGN AND EVALUATION PROCESS OUTLINE (DEPO)



“NEVER SPEND TIME DOING BY HAND WHAT YOU CAN AUTOMATE WITH A COMPUTER”
(DR. PEVEY – UNIVERSITY OF TENNESSEE KNOXVILLE)

LEARN BY PLAYING: REINFORCEMENT LEARNING



REINFORCEMENT LEARNING HAS SEVERAL ADVANTAGES OVER TRADITIONAL METHODS

- Reward structure can be changed to prioritize different goals (i.e., cost, physical footprint, etc)
- Agents can dynamically explore environments in real-time and react optimally under different conditions (e.g., before/after detection)
- Reinforcement learning can explore different states of knowledge of adversaries (e.g., varying knowledge of facility layout)

Project Goal: Develop designer and adversarial agents for physical protection systems

- Designer agent will propose candidate PPS layouts depending on user-set criteria
- Adversarial agent will find optimal paths and severe vulnerabilities

PPO IS SO YESTERDAY

- Initially tried both proximal policy optimization (PPO) and option-critic for the adversarial and planning agent respectively
- Didn't have a lot of success so opted to leverage the SOTA Dreamer family of models
- These approaches utilize world models (recurrent state space models)
- Instead of the actor/critic(s) learning directly from the environment, Dreamer algorithms learn from imagined rollouts from the RSSM
- Dreamer algorithms are more sample efficient and less brittle than past RL algorithms, but harder to implement

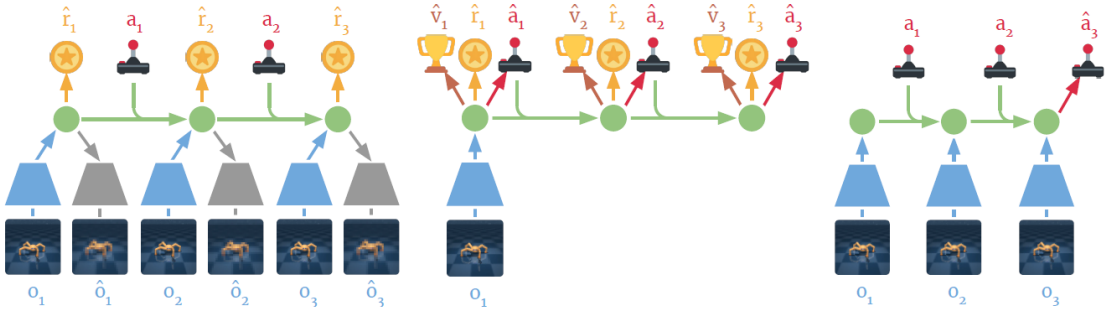


Image credit: Hafner et al.

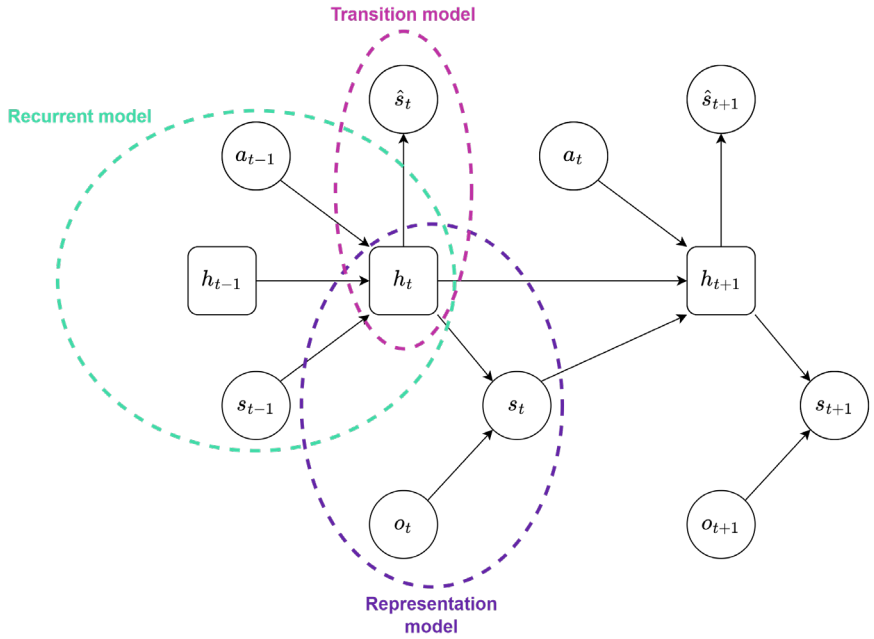
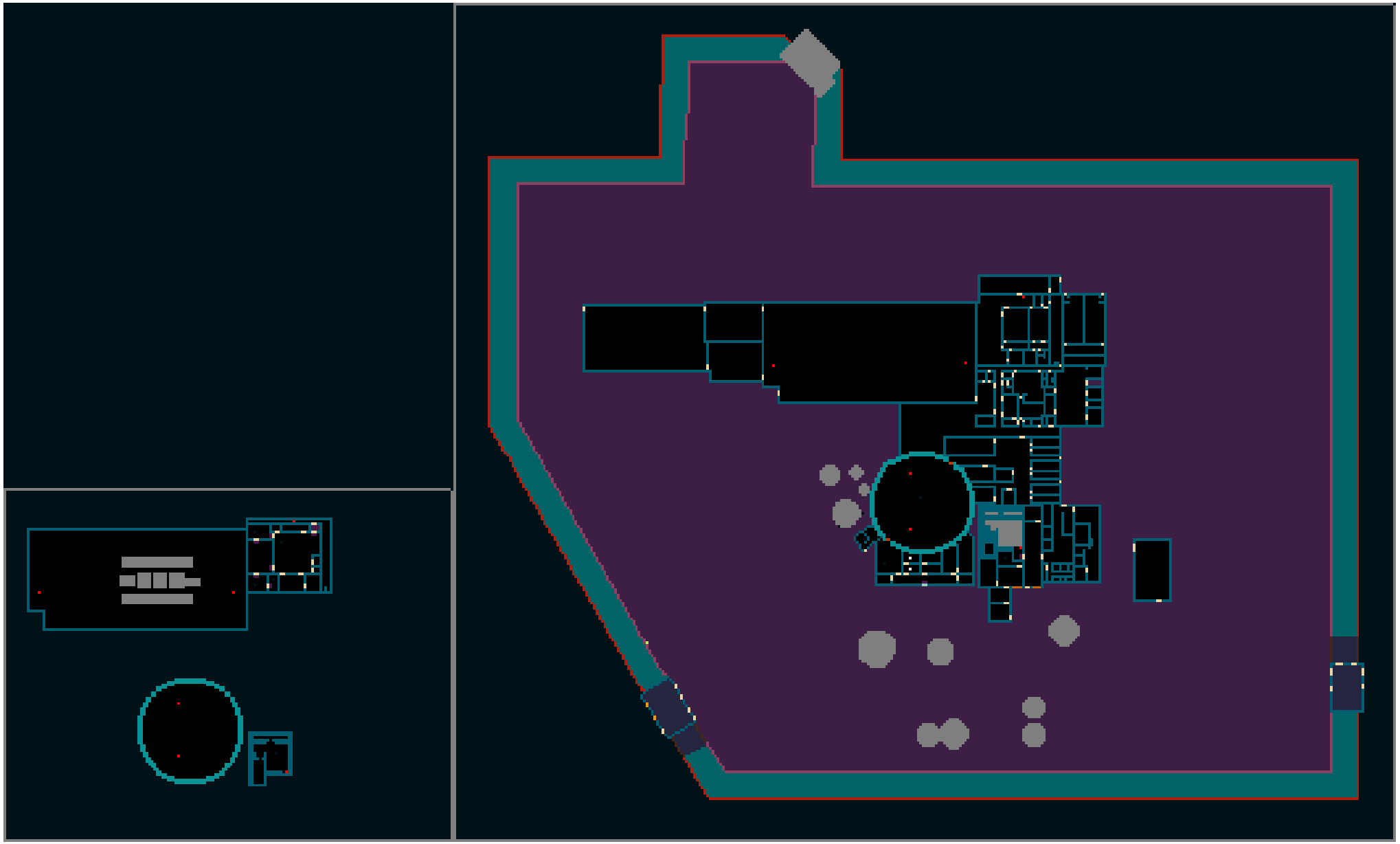


Image credit: Michele Milesi/eclecticsheep

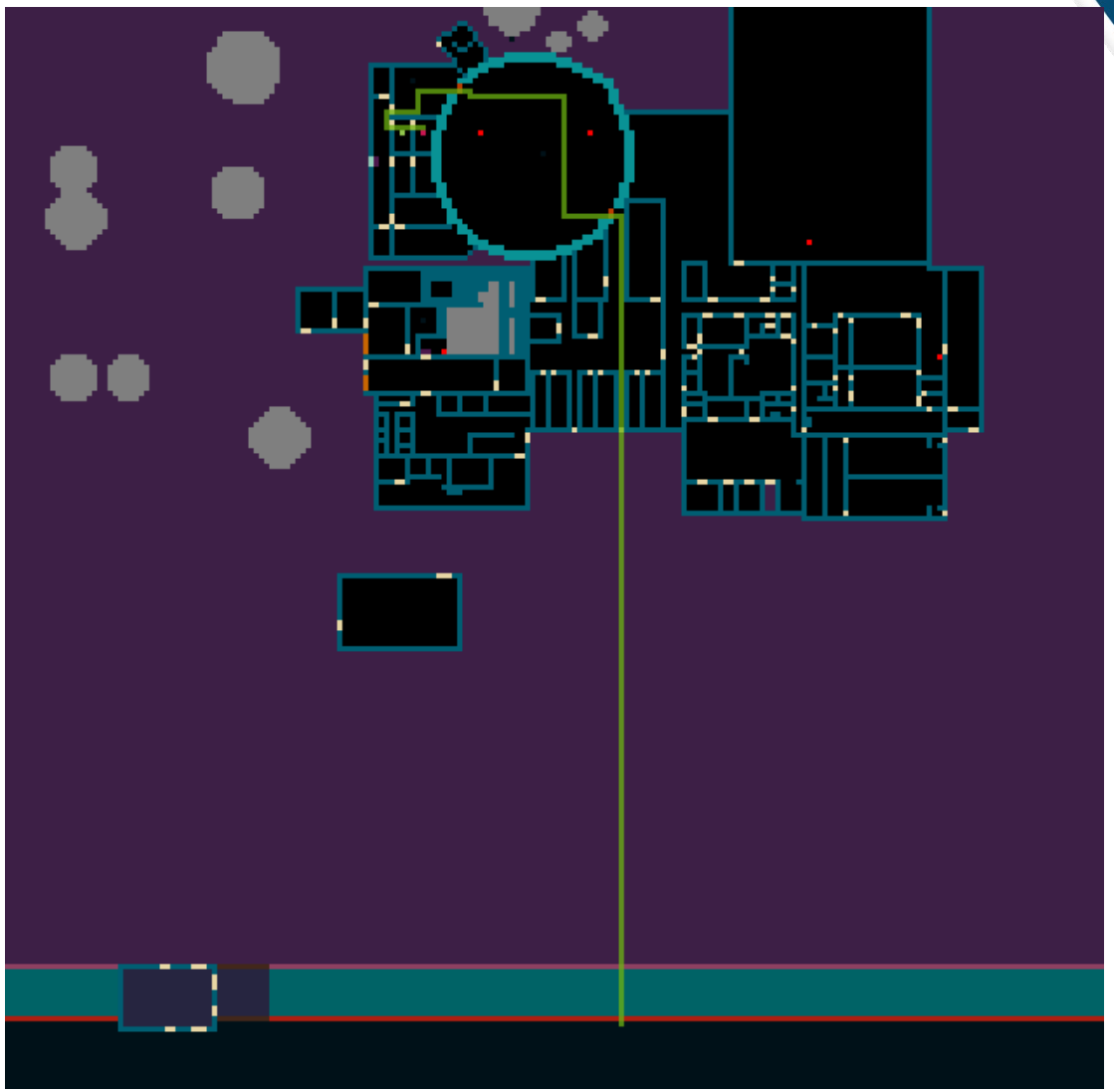
ADVERSARIAL AGENT (DV3)

SO WHERE DO WE START? THE ENVIRONMENT

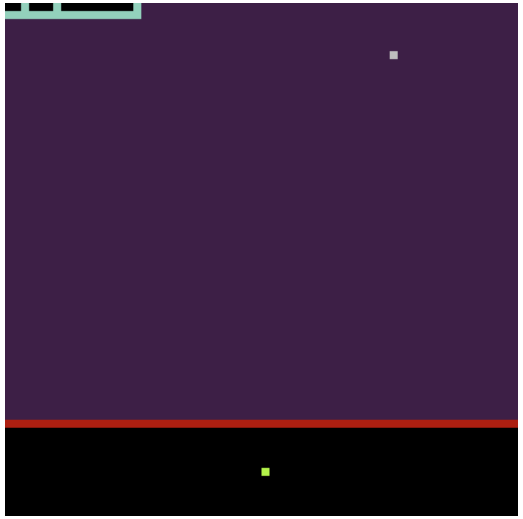


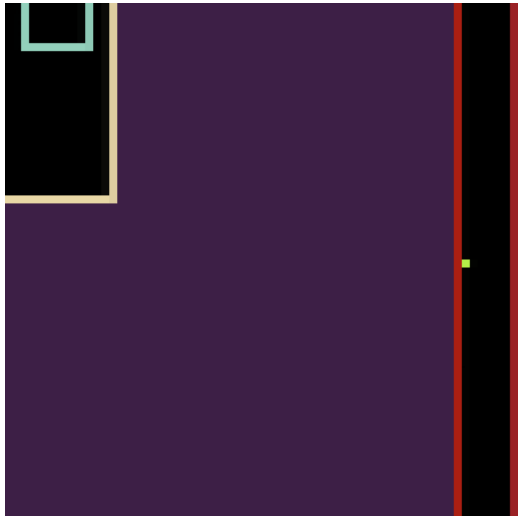
WHAT'S THE OBJECTIVE? THE REWARD SPACE

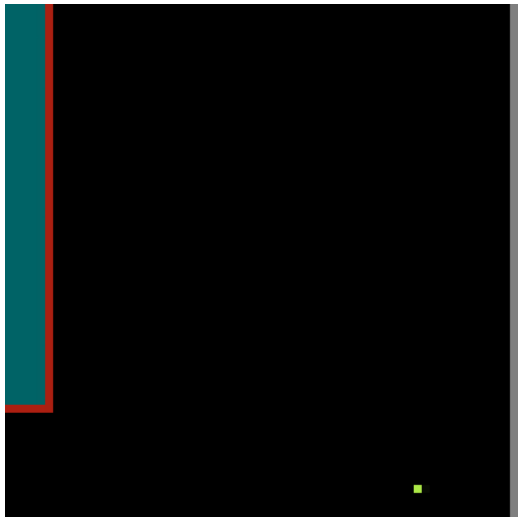
- Collect as much value as possible (obviously)
- Specific objective varies by map
- Generally, agent is tasked with reaching objective(s) without being intercepted
- Agent needs to learn the most effective paths
- In some instances, secondary targets are available

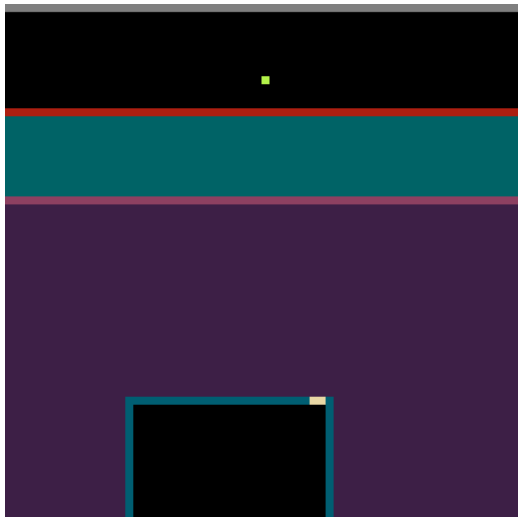


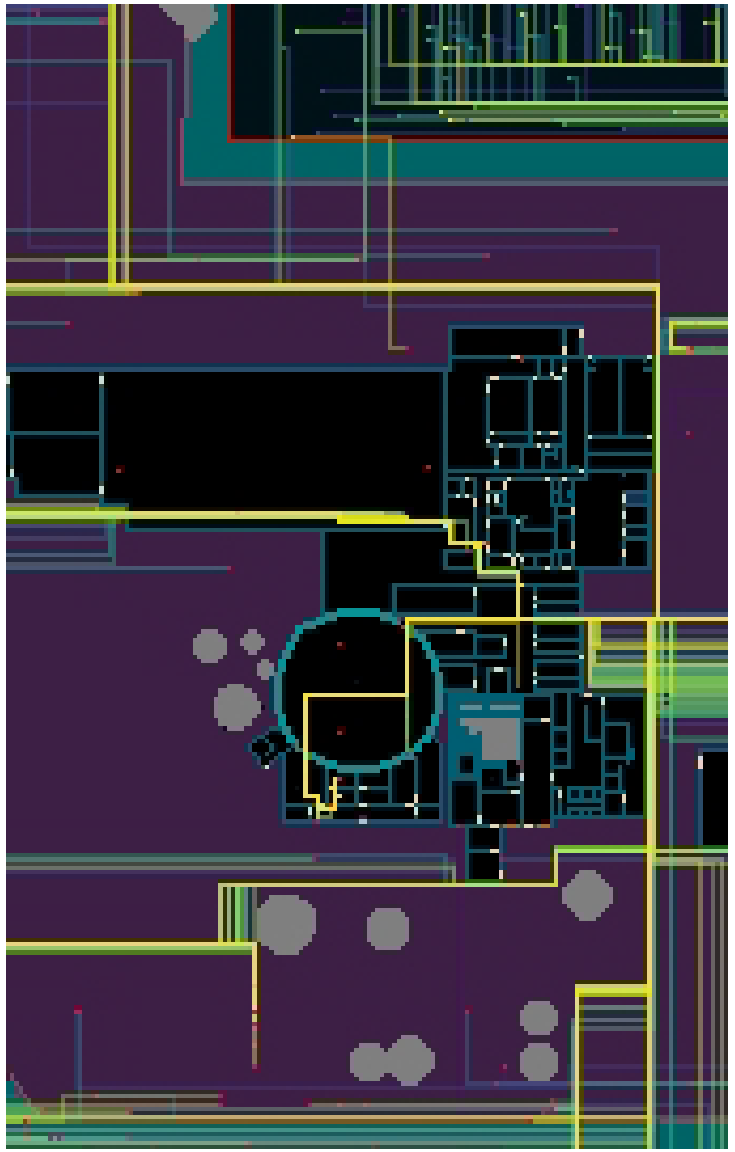
ADVERSARIAL AGENT CAN LEARN THE MOST VULNERABLE PATH ENTIRELY THROUGH SELF PLAY











AGENTS ADAPT TO CONDITIONS IN REAL-TIME (PENALTY AVOIDANCE)

- A variety of emergent actions can be seen when an agent is detected
  - Reusing destroyed barriers to leave a facility
  - Attempting to leave if discovered
  - Secondary target fall back acquisition
- Agents exhibit maximally destructive behavior; secondary targets will be attempted wherever possible

PLANNER AGENT (DIRECTOR-DV3)

VERY INTERESTING, BUT CAN WE GET THE COMPUTER TO DO ITERATIVE DESIGN?

- Design is a much more difficult task
- Not simple to implement, less “out-of-the-box” than adversarial
- Fewer literature examples
- **Goal**: leverage HRL with a worker/manager scheme to perform “auto complete” on parts of the design with human-in-the-loop
- Wasted some considerable time with option-critic (”easier” to implement)
- Working with a modified Director architecture
  - Manager/worker architecture with shared RSSM + goal encoder
  - Custom encoder/decoders
  - 20+ neural networks concurrently trained
- Director is nice w/ world model and single reward

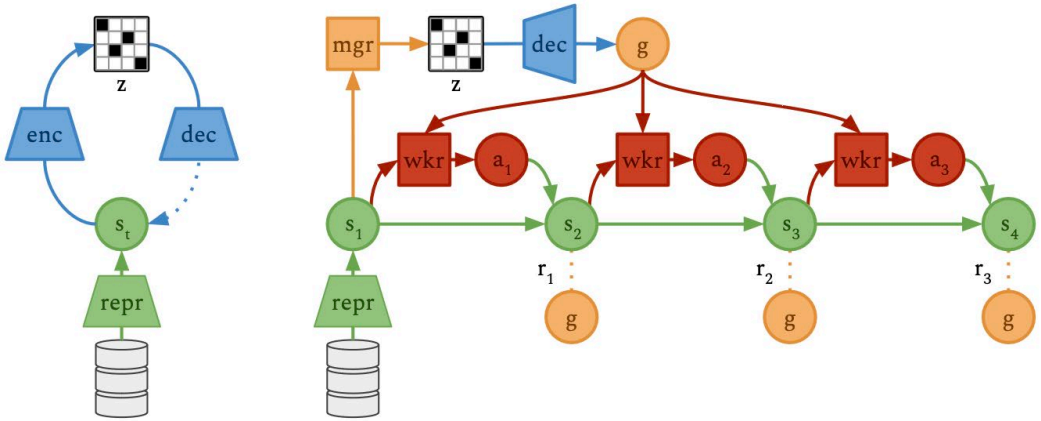


Image credit: Hafner et al.

SURELY THE ACTION SPACE FOR DESIGN IS MORE COMPLEX (ACTION)?

- Action space is much larger and complex
- Agent can change any tile to one of 22+ different objects
- Action space rapidly gets out of hand
  - Somewhat mitigated through a shared worker trunk with triple action head
- Procedural Content Generation via RL (PCGRL) provides some candidate
  - Narrow: agent is presented (x, y) and must select a tile to place/change
  - Turtle: agent exists in world, actions change tile (must traverse)
  - **Wide**: agent selects (x,y) and tile

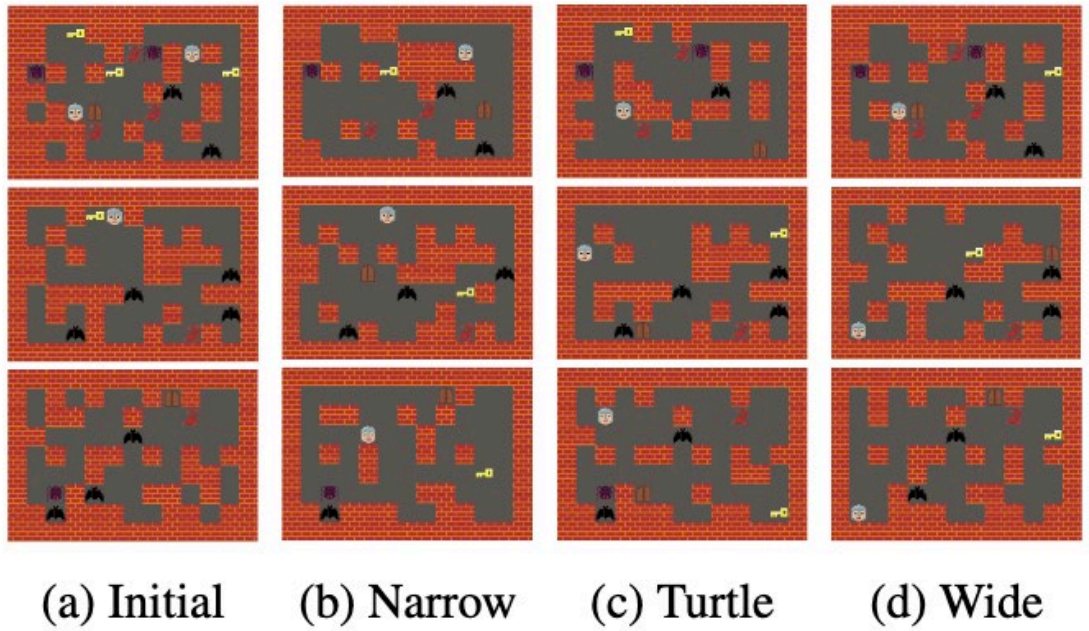
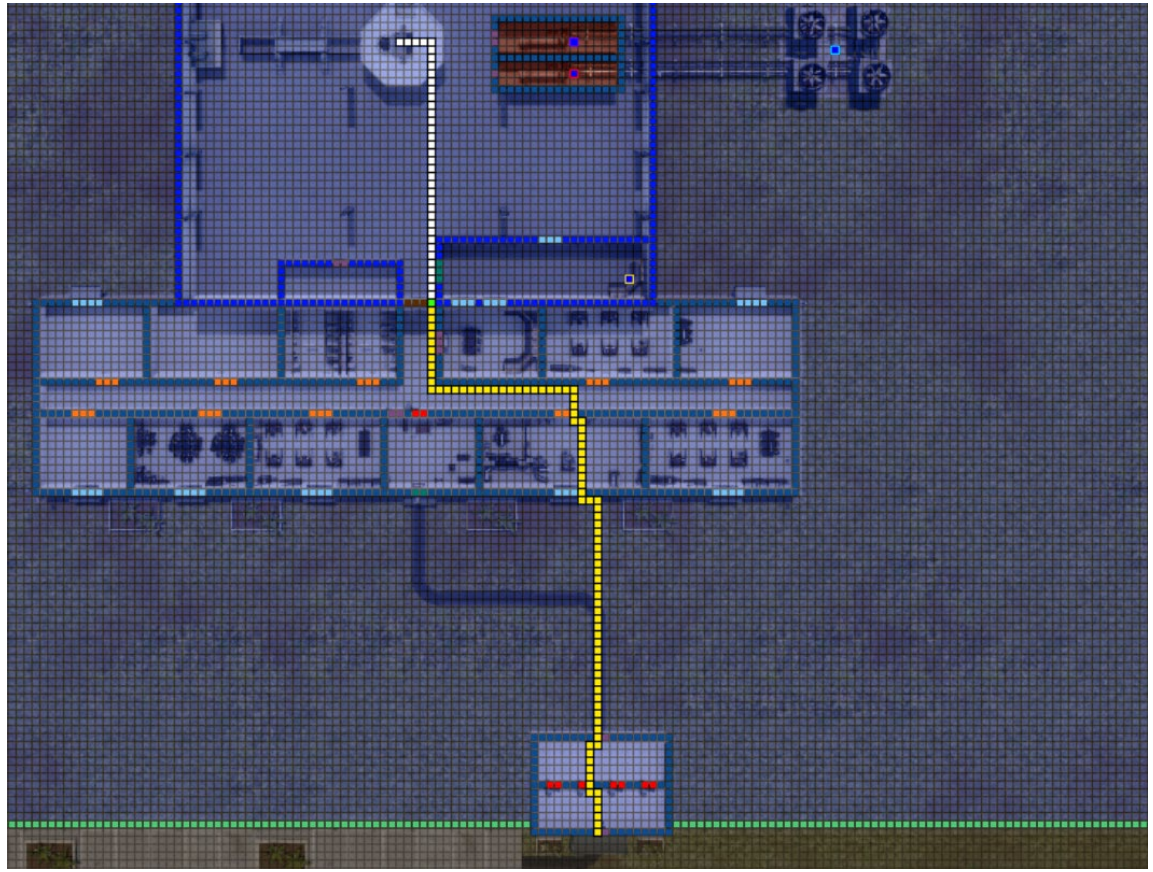


Image credit: Khalifa et al.

HOW DO WE EVEN ASSESS THE “GOODNESS” OF A LAYOUT (ENVIRONMENT)?

- DEPO is an iterative process with multiple analyses common metrics include:
  - Probability of Interruption
  - Probability of Detection
  - Critical Detection Point
  - Adversary Sequence Diagrams
- This work focuses on probability of interruption
- A good layout balances (approximate) cost, footprint, performance metrics, and emergency requirements
- Initially we are focusing on the balance between performance and (approximate) cost

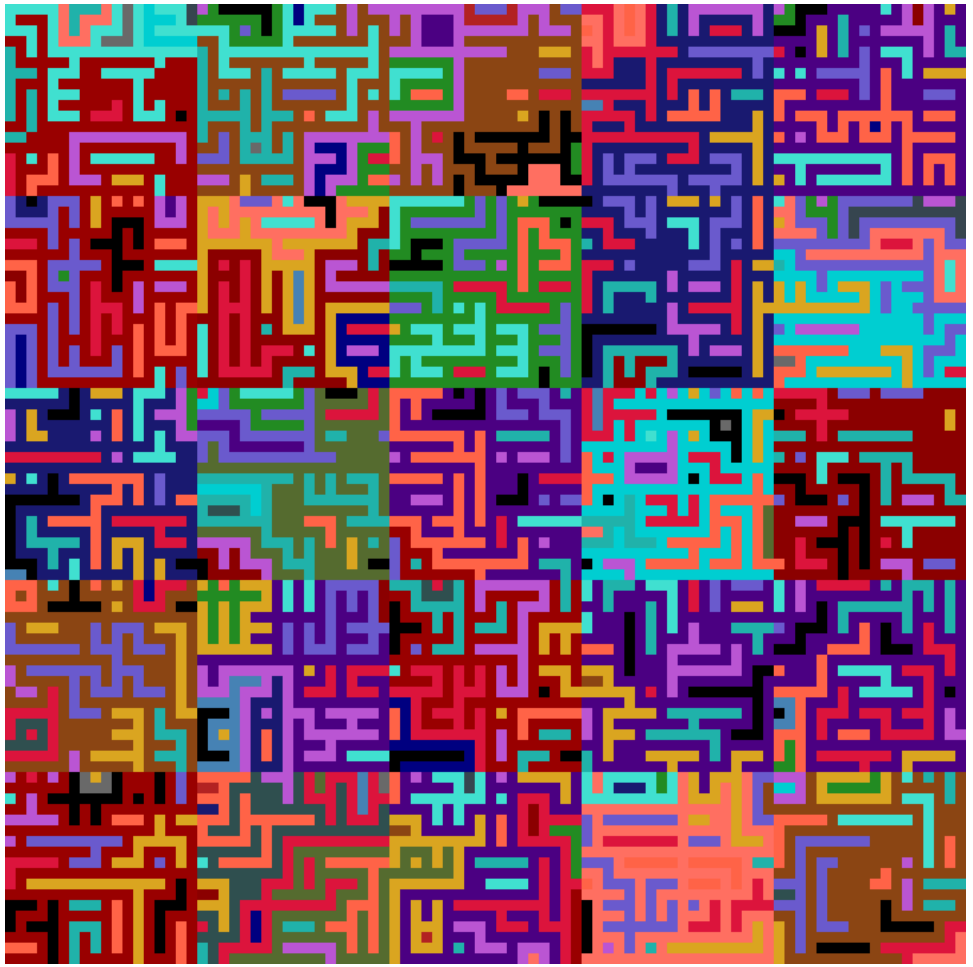


HOW DO WE CALCULATE INCREMENTAL ACTIONS (REWARD)?

- Current state of practice: Run Dijkstra's algorithm on a weighted graph, find the cheapest path
- How can we determine the value of a single tile change in the context of tools usage and probability of detection?
  - If there's a wall instead of a sensor here, what value does that have? Both delay *and* probability of detection are important
- Solution: Maintain two separate graph representations
  - First graph represents the cost *after* detection (so functionally the smallest delay possible for each item type)
  - Second graph is the real, "working" graph calculates the cost if detected (from graph 1) times the probability of detection
  - Both graphs are run and updated when agent changes a tile

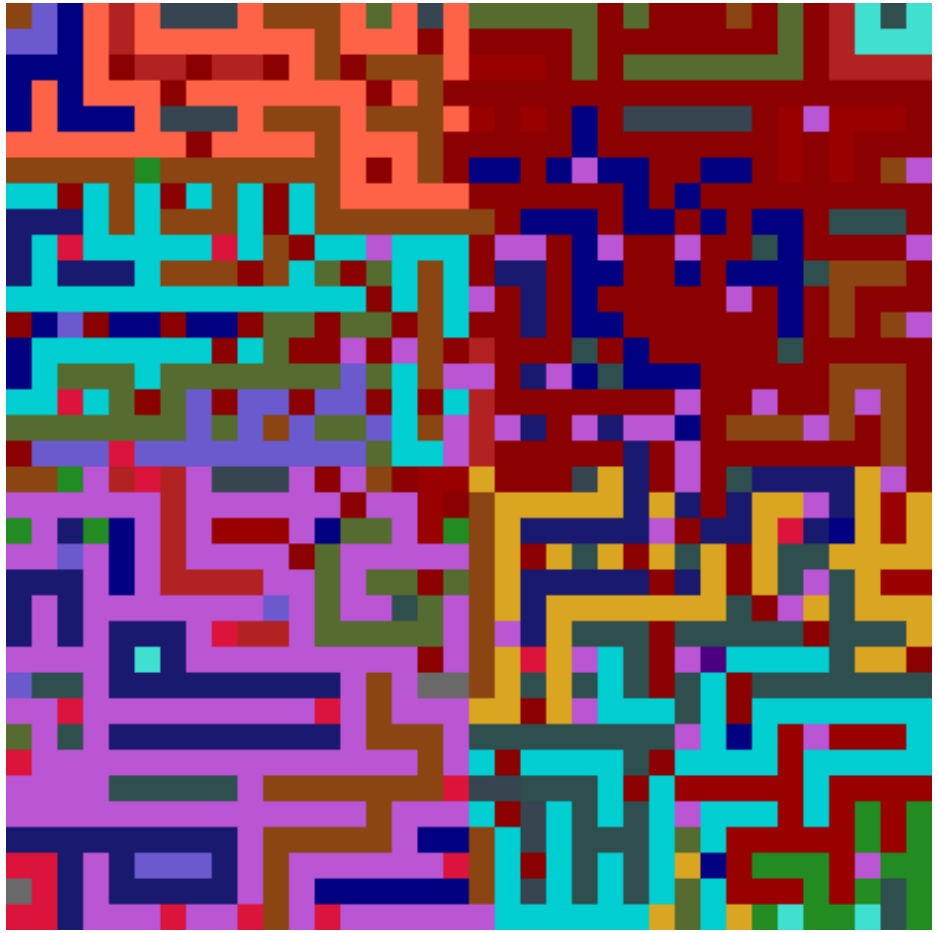
HOW DO WE BALANCE DELAY AND COST?

- $r = S * \left(\frac{\Delta(\text{pathlength})}{\text{scale}_1} - \frac{\varepsilon\Delta(\text{cost})}{\text{scale}_2} \right)$
- Path length is averaged over the entire maze to ensure a dense reward (every edit does *something*)
- Costs informed by real world metrics (fences cost less than reinforced steel doors)
- Accelerate learning by randomly generating mazes (sampled from a pretrained VAE)
- Limiting edits to about 30% of tiles
- Future: Manager edit gate to learn when to stop editing



AI AGENTS CAN SUCCESSFULLY FUNCTION AS AN “AUTO-COMPLETE” FOR PHYSICAL SECURITY DESIGN

- Randomized layouts are provided as input
- Agent tasked with maximizing “path length” while balancing cost
- Agent can access 22 different physical security elements
- “Pan-like” observation space could be improved
- Demonstrated on smaller scale problem (~5x smaller than full facility)
- Future: Patch-like rotation? How could this be integrated with Director’s continuous observation space expectation



CONTINUING WORK AND OTHER INTERESTING PROJECTS 🤖

- 🚀 Ongoing: adversarial agent capabilities beyond state-of-practice
  - 🕶️ Partial observability
  - 🏔️ Tool-specific traversal changes
  - 📍 Region-based targets
  - 🔄 Human-in-the-loop evaluations
  - 🌟 … and more!
- 🚡 Ongoing: planning agent
  - 🧪 Completing a basic toy-level demonstration (full ruleset)
  - 📈 Scaling up to larger maze sizes
  - ⚓ Adding “immutable” objects

ACKNOWLEDGEMENTS

This work was funded by the Department of Energy, Office of Nuclear Energy's Advanced Reactor and Safeguards (ARSS) program.

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.

QUESTIONS?