Open World Planning for Robots via Hindsight Optimization

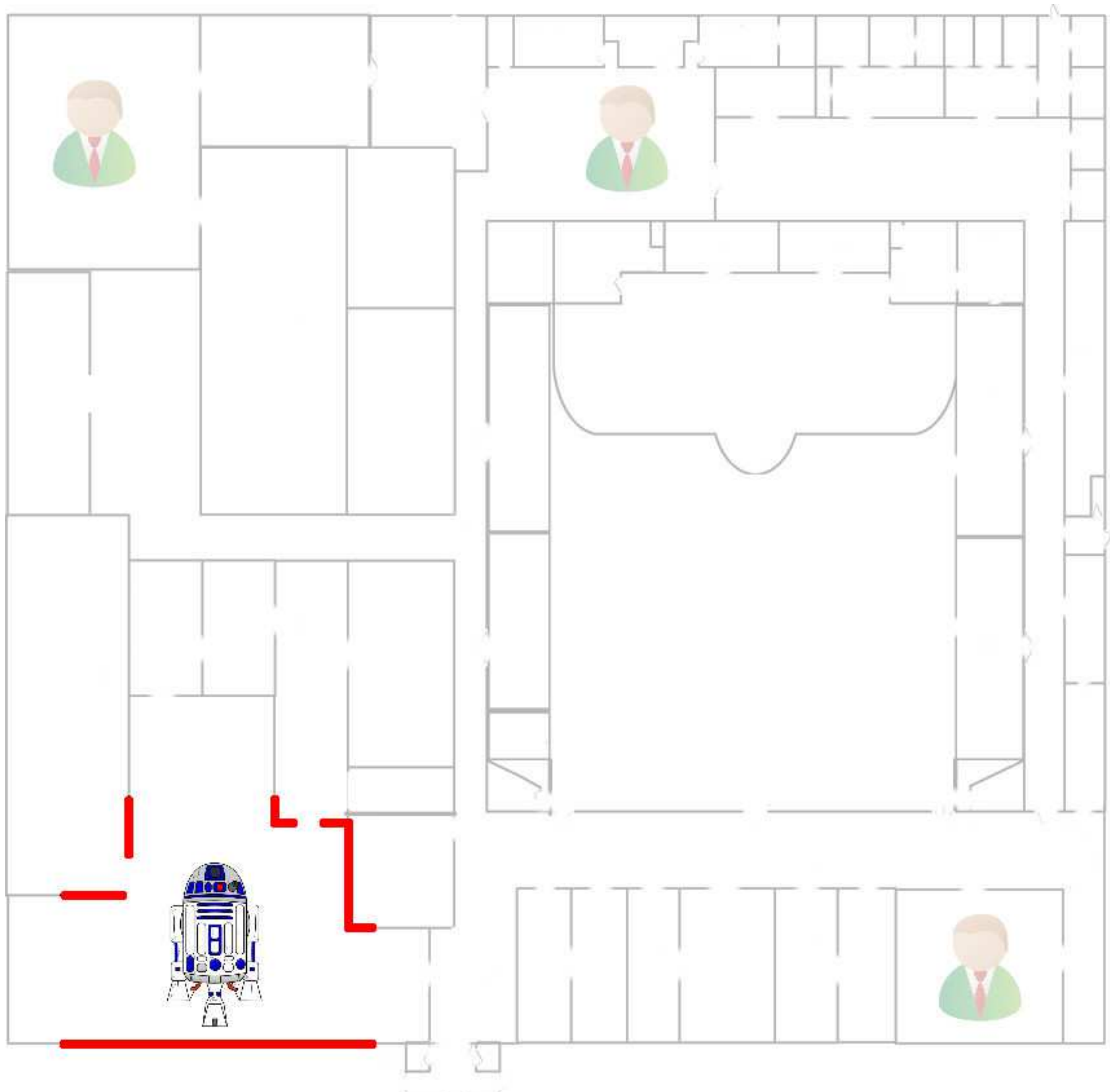
Scott Kiesel[1], Ethan Burns[1], Wheeler Ruml[1], J. Benton[2], Frank Kreimendahl[1]

University of New Hampshire [1]

SIFT Smart Information Flow Technologies [2]

We are grateful for funding from the DARPA CSSG program (grant H R0011-09-1-0021) and NSF (grant IIS-0812141).

Open World Planning - Search and Rescue



Search and Rescue Domain

- Robot agent
- Unknown building/map layout
- Unknown victim locations
- Unknown number of victims
- Search time limit

Previous Approaches

- Talamadupula et al. (ICAPS '09, AAAI '10, TIST '10)
  - ad-hoc assumption: $roomExists(x) \rightarrow personExistsIn(x)$
- Joshi et al. (ICRA '12)
  - based on FODD approximations
  - hours of offline planning
- Optimization in Hindsight with Open Worlds (OH-wOW)
  - general
  - principled
  - easy to implement (and extend)

Hindsight Optimization

Select action that maximizes expected reward.

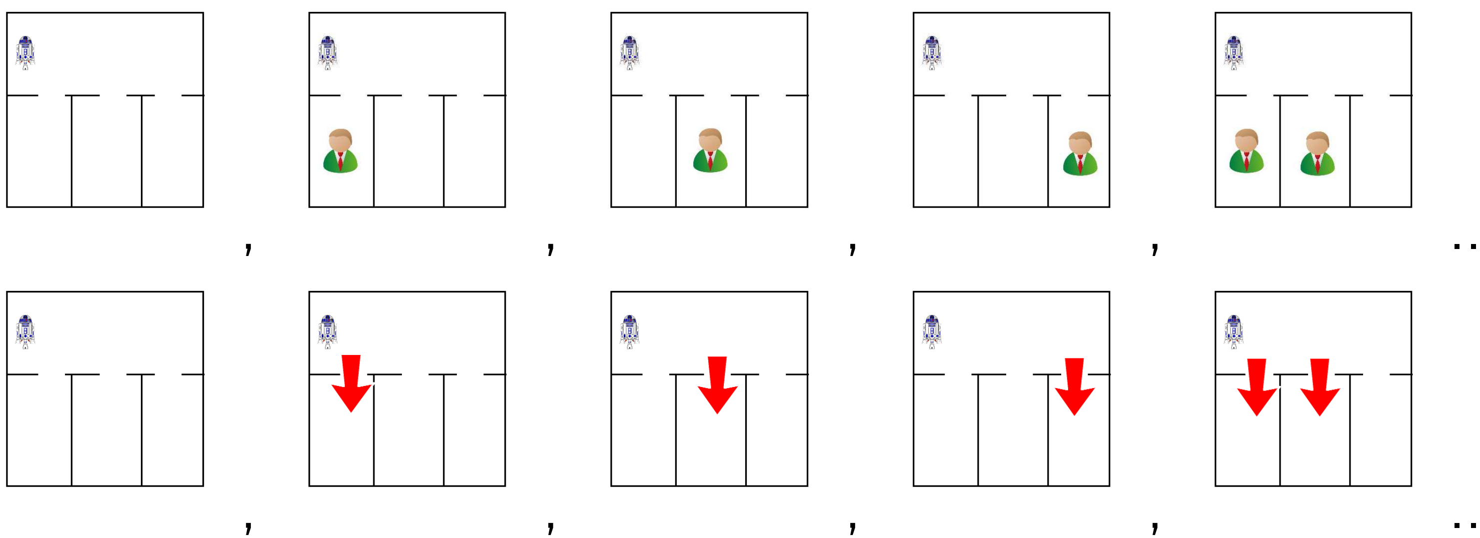
reward = cumulative reward following optimal plan

Hindsight Optimization

Select action leading to states with highest expected reward.

reward = reward of plan out of all possible plans with best average reward over all configurations

$$V^*(s_1) = \min_{A=\langle a_1,\ldots,a_{|A|}\rangle} \mathop{E}_{\langle s_2,\ldots,s_{|A|}\rangle} \left[\sum_{i=1}^{|A|} R(s_i, a_i) \right]$$

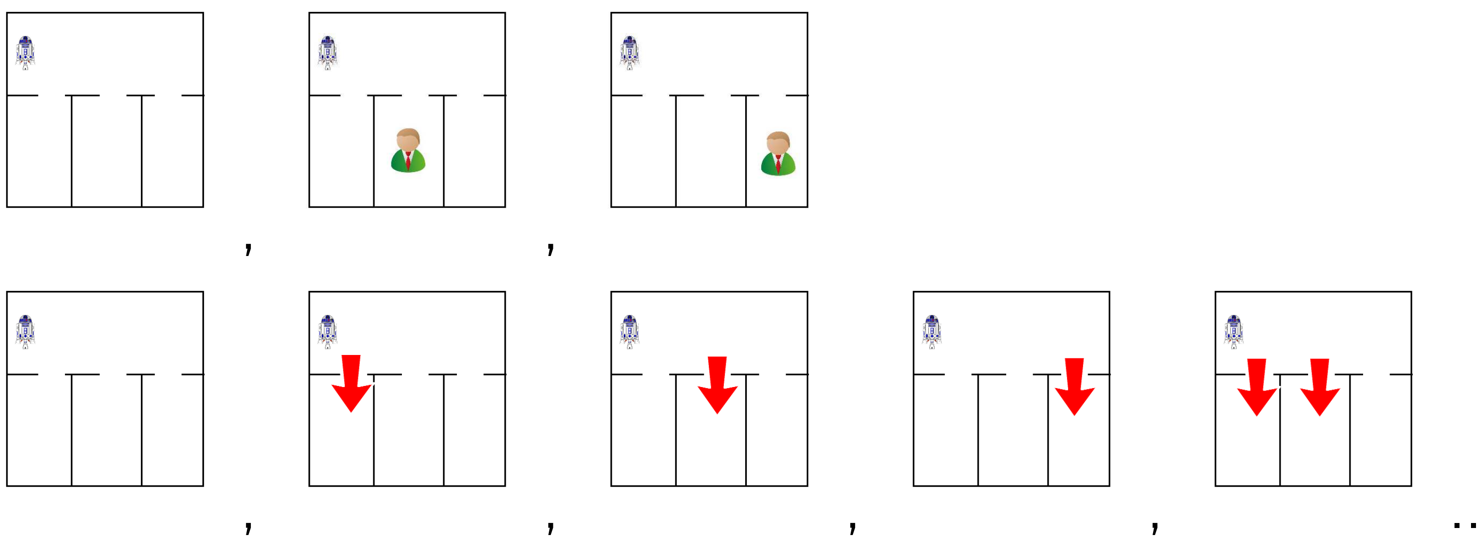


Hindsight Optimization

Select action leading to states with highest expected reward.

reward ≈ reward of plan out of all possible plans with best average reward across sampled configurations

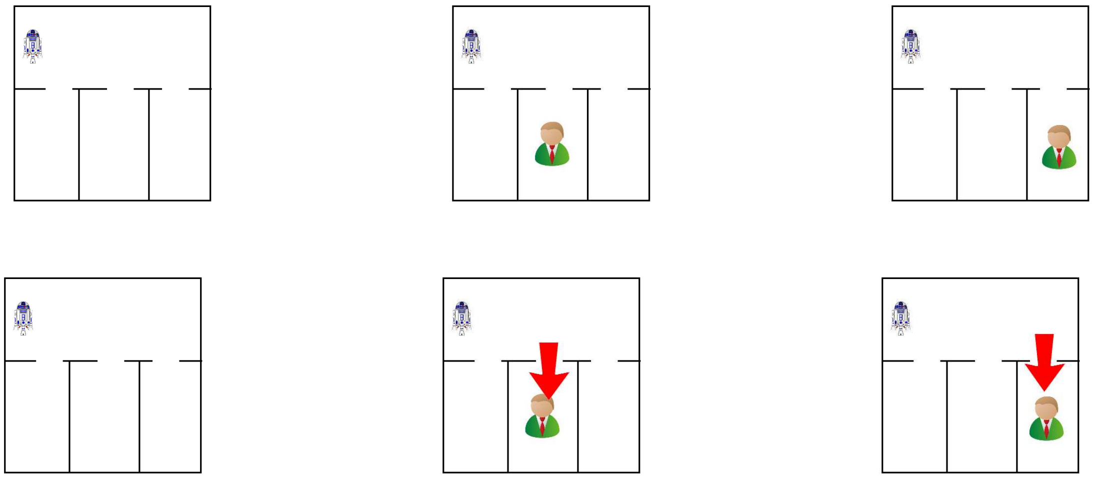
$$\hat{V}(s_1) = \min_{A=\langle a_1,\ldots,a_{|A|}\rangle} \underset{\langle s_2,\ldots,s_{|A|}\rangle}{E} \left[\sum_{i=1}^{|A|} R(s_i, a_i)\right]$$



Select action leading to states with highest expected reward.

reward ≈ average reward of best plan in each sampled configuration

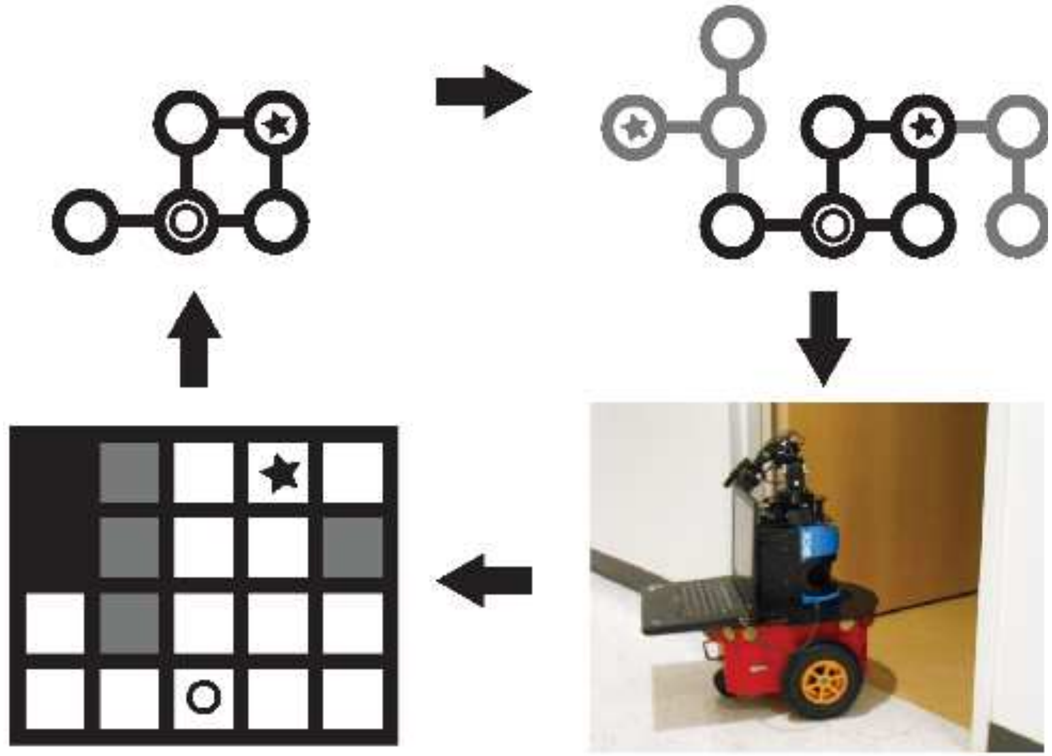
$$\hat{V}(s_1) = \underset{\langle s_2, s_3, \ldots \rangle}{E} \left[\min_{A=\langle a_1, \ldots, a_{|A|} \rangle} \sum_{i=1}^{|A|} R(s_i, a_i) \right]$$



Optimization in Hindsight with Open Worlds

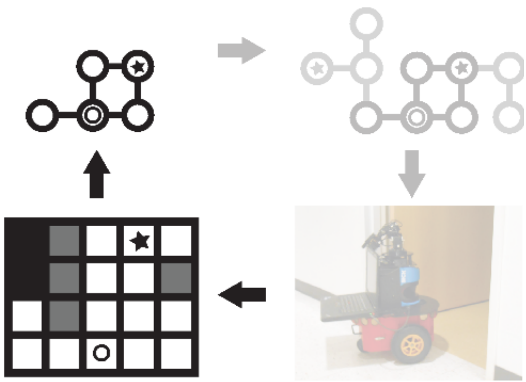
OH-wOW Implementation for Search and Rescue

1. Sense
2. Sample
3. Plan
4. Act



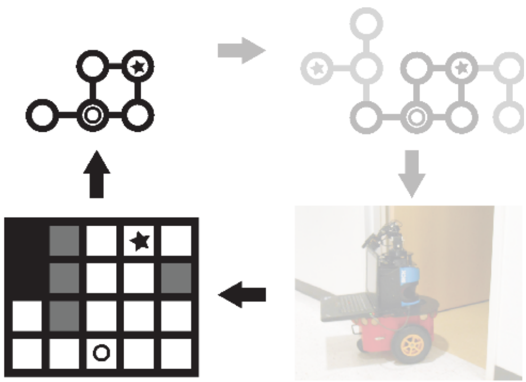
Sensing and Observations

1. Sense
2. Sample
3. Plan
4. Act

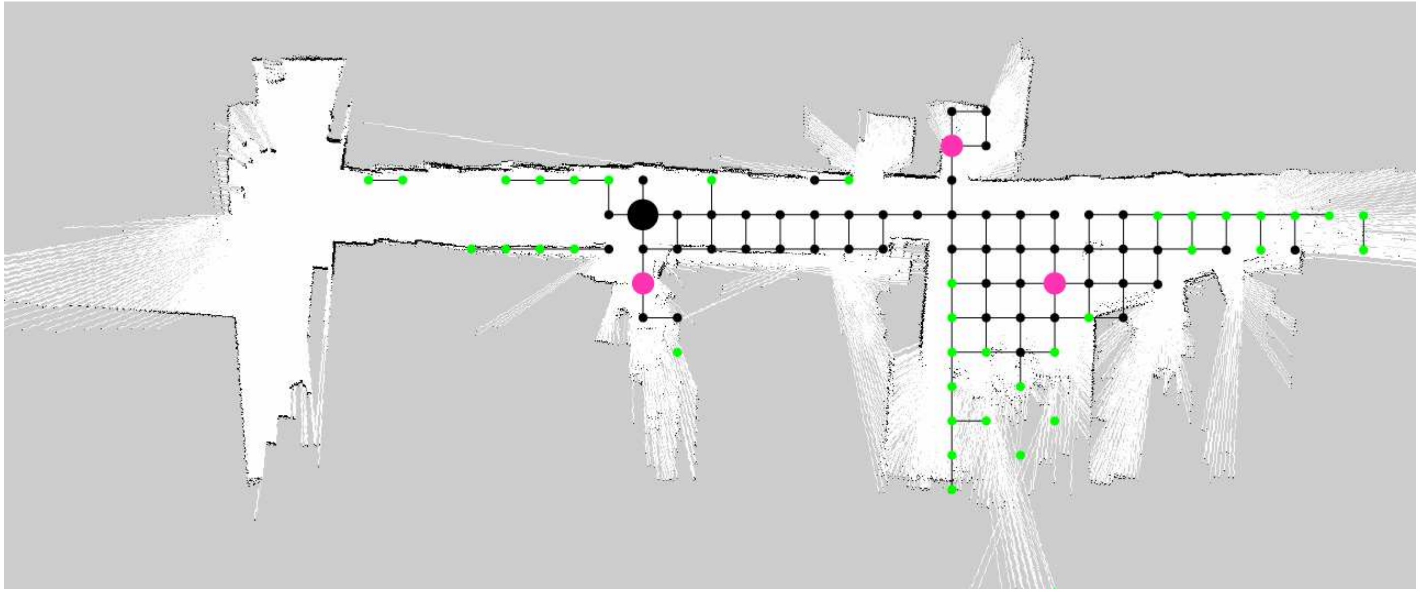


- SLAM (ROS gmapping)
  laser rangefinder
- Topological Map
  rough construction
- Person Detector

1. Sense
2. Sample
3. Plan
4. Act

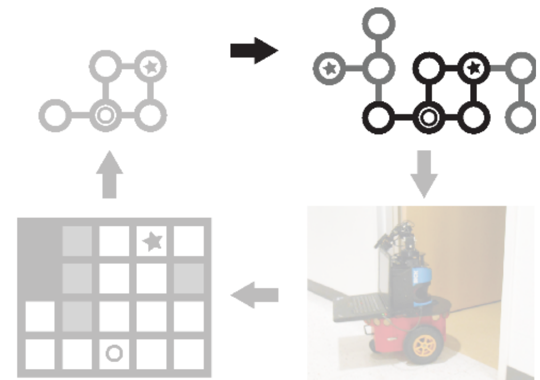


Sensed Occupancy Grid with Topological Graph Overlayed



Sampling Possible Worlds

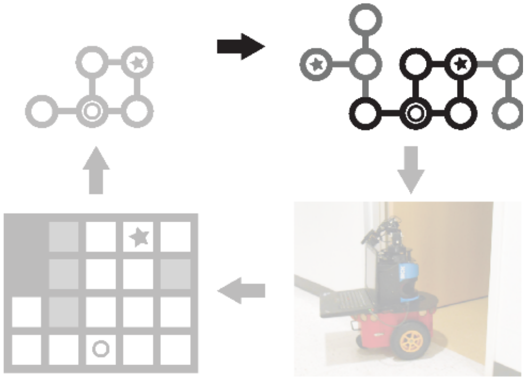
1. Sense
2. Sample
3. Plan
4. Act



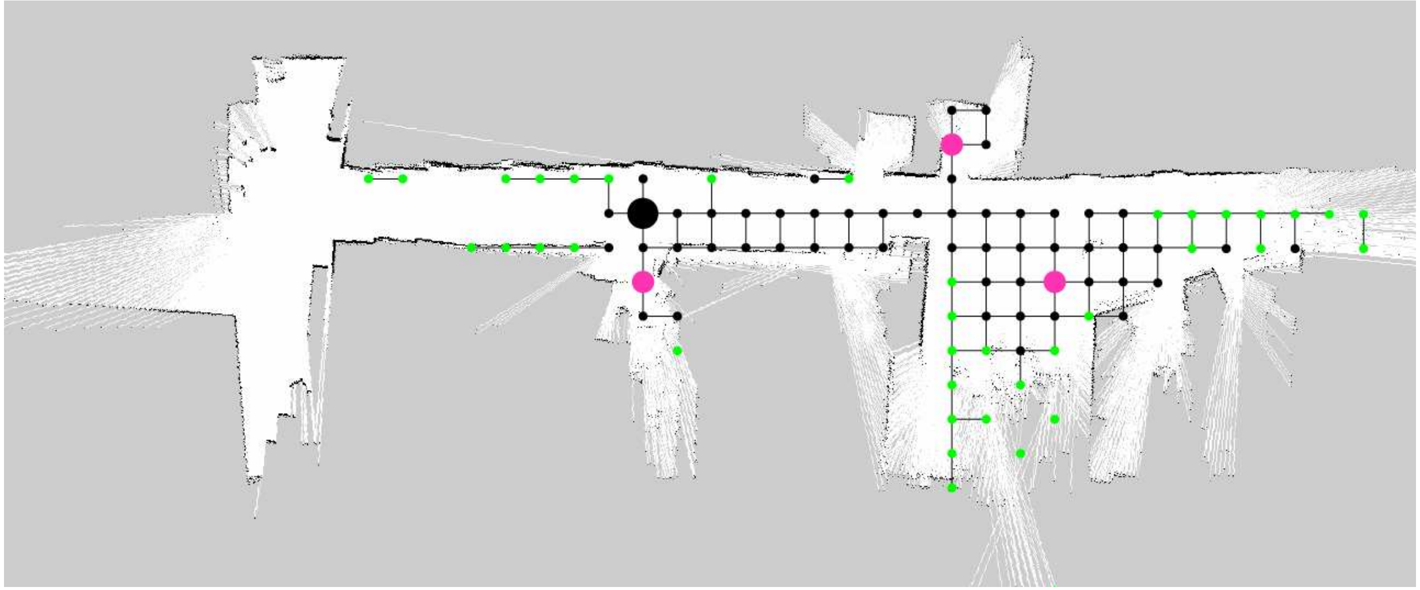
- Current Knowledge
  - observed
  - known to be *true*
- Expectation
  - prior domain knowledge
  - bias

Sampling Possible Worlds

1. Sense
2. Sample
3. Plan
4. Act

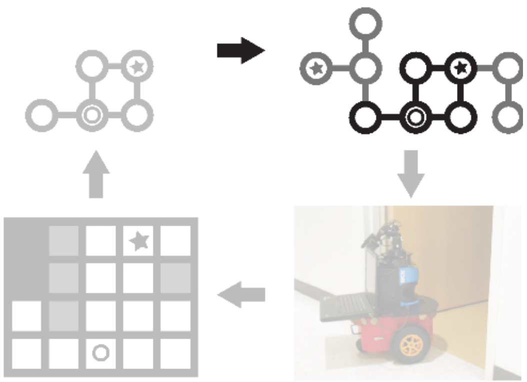


Known Partial World State

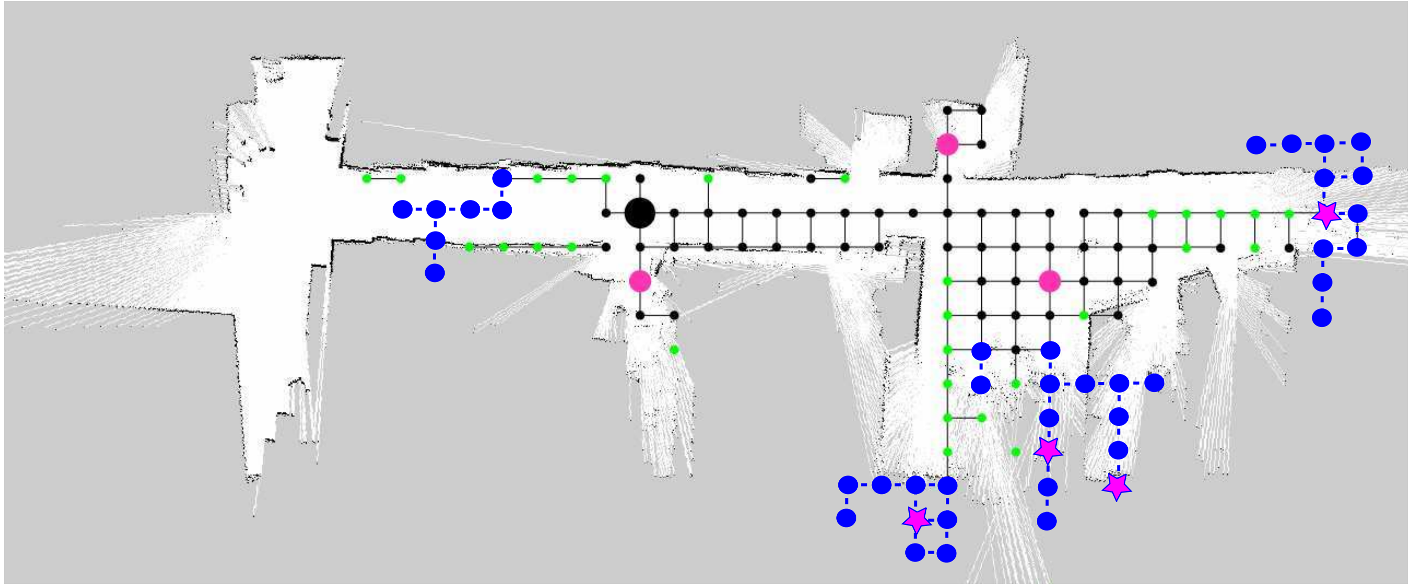


Sampling Possible Worlds

1. Sense
2. Sample
3. Plan
4. Act

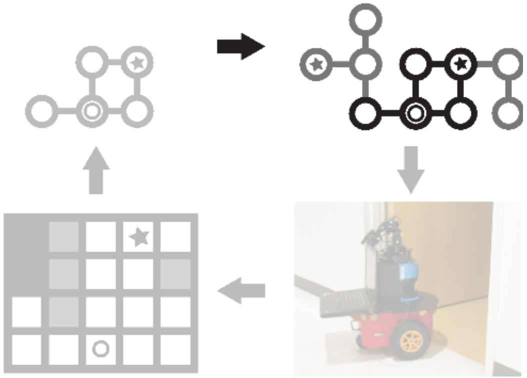


Sampled "Complete" World State



Planning in Sampled Worlds

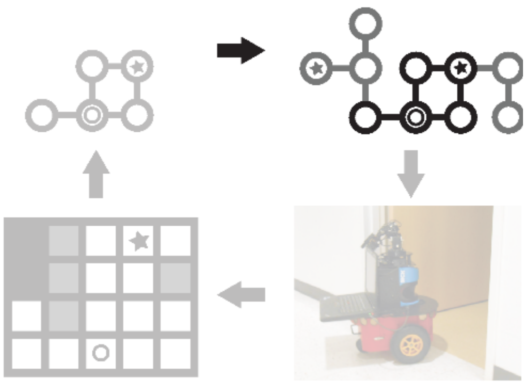
1. Sense
2. Sample
3. Plan
4. Act



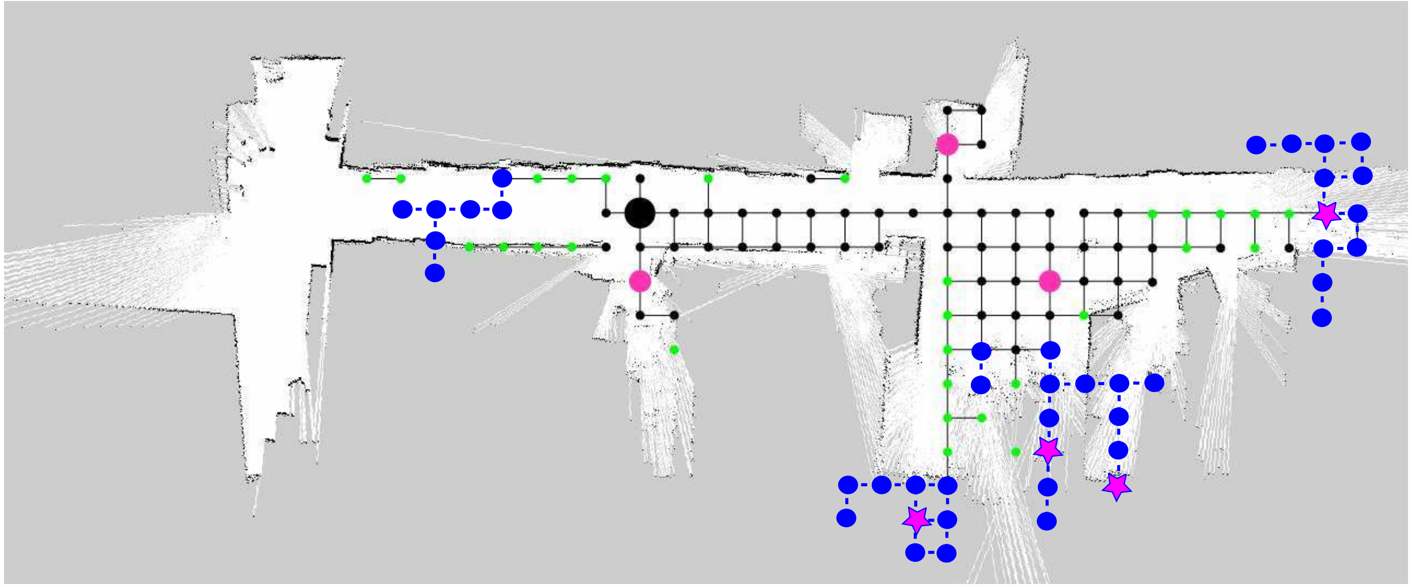
- Fully Known
- Deterministic
- Classical Planners or
- Domain Specific Planners

Planning in Sampled Worlds

1. Sense
2. Sample
3. Plan
4. Act

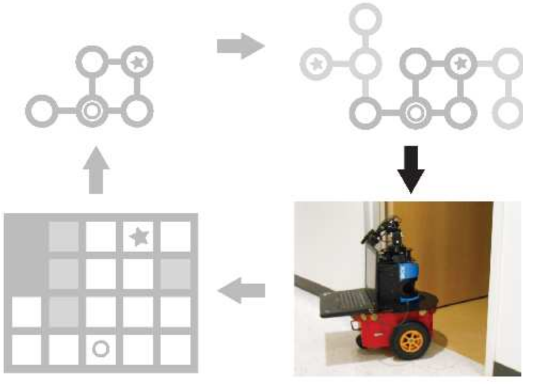


A Single Sample



Acting in Sampled Worlds

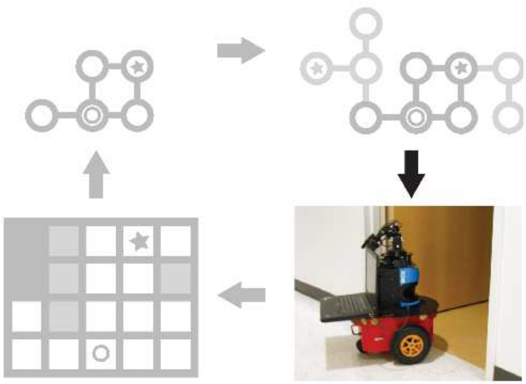
1. Sense
2. Sample
3. Plan
4. Act

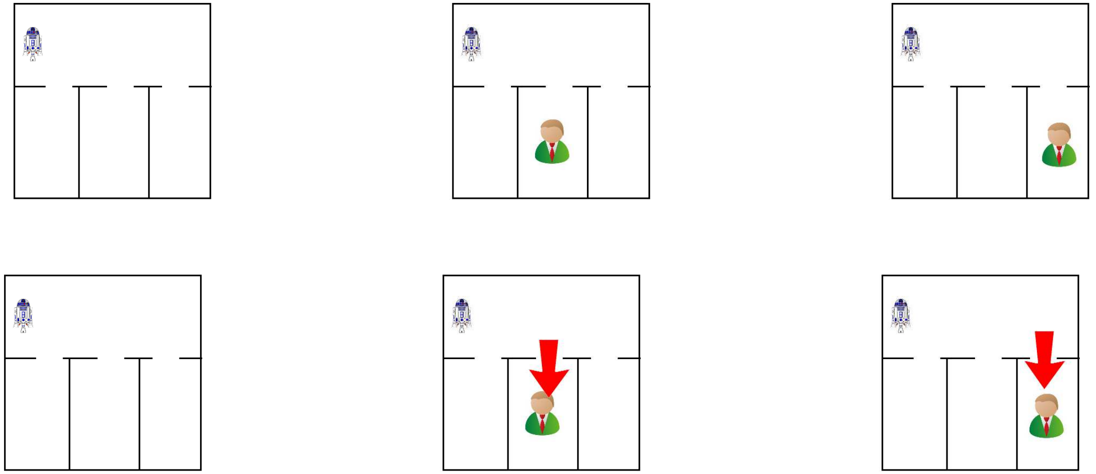


- Execute Best Currently Available Action
  maximize expected reward

Acting in Sampled Worlds

1. Sense
2. Sample
3. Plan
4. Act





Results

Search and Rescue

UNH CS Offices, Pioneer 3-DX, SICK LMS500, ROS Fuerte

	victims found			
deadline	0	1	2	3
1 minute	4	6	0	0
5 minutes	0	7	3	0
10 minutes	0	3	4	3

Joshi et al:

4 hours precomputation, 3 victims

constant time table lookup

OH-wOW:

no precomputation

0.18 sec avg max step time, 3 victims (256 samples)

2.7 sec avg max step time, 10 victims (256 samples)

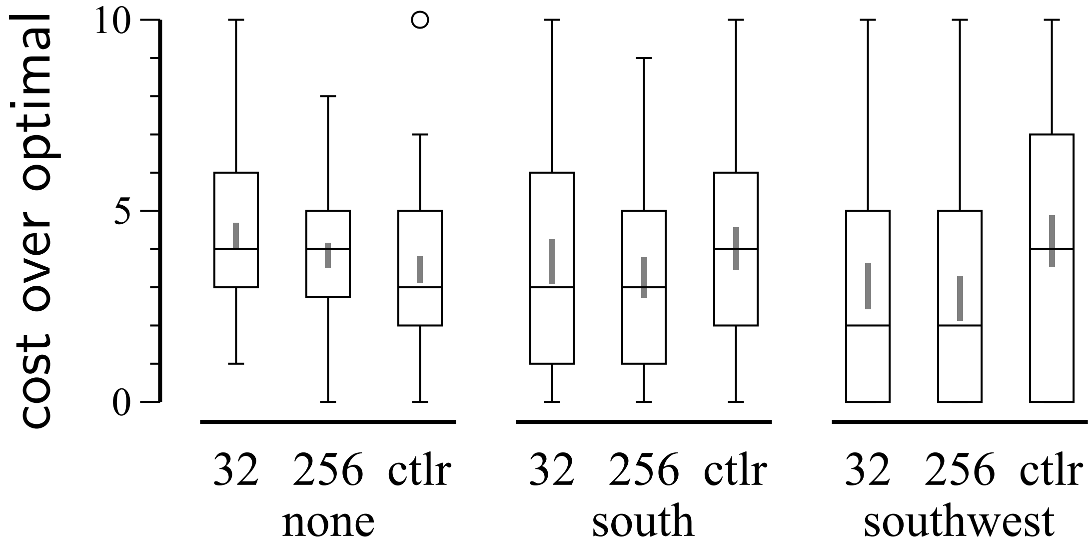
UNH CS Offices, Pioneer 3-DX, SICK LMS500, ROS Fuerte

	victims found			
deadline	0	1	2	3
1 minute	4	6	0	0
5 minutes	0	7	3	0
10 minutes	0	3	4	3

OH-wOW:

- is online,
- computes the next action quickly,
- and handles the tradeoff between hard and soft goals.

Search and Rescue in Simulation



OH-wOW:

- leverages domain specific knowledge,
- and can beat a handcoded controller.

Omelette Domain in Simulation

Levesque (AAAI '96)

	planning time (seconds)			
	3 eggs	step	4 eggs	step
Bonet et al (IJCAI '01)	185	-	-	-
Levesque (IJCAI '05))	1.4	-	1,681	-
OH-wOW	12.9	0.52	76.7	1.57

Levesque plans are longer than OH-wOW

OH-wOW:

- is online,
- computes the next action quickly,
- and finds cheaper cost solutions.

Conclusion

Limitations

- Scalability of the underlying planner
  - leverage large body of literature
- Calls underlying planner repetitively
  - embarassingly parallel
- Vulnerable to black swans during sampling
  - importance sampling
- Regenerates world samples at every step
  - reuse samples until world "changes"

(see Yoon et al. ICAPS '10 for HO Optimizations)

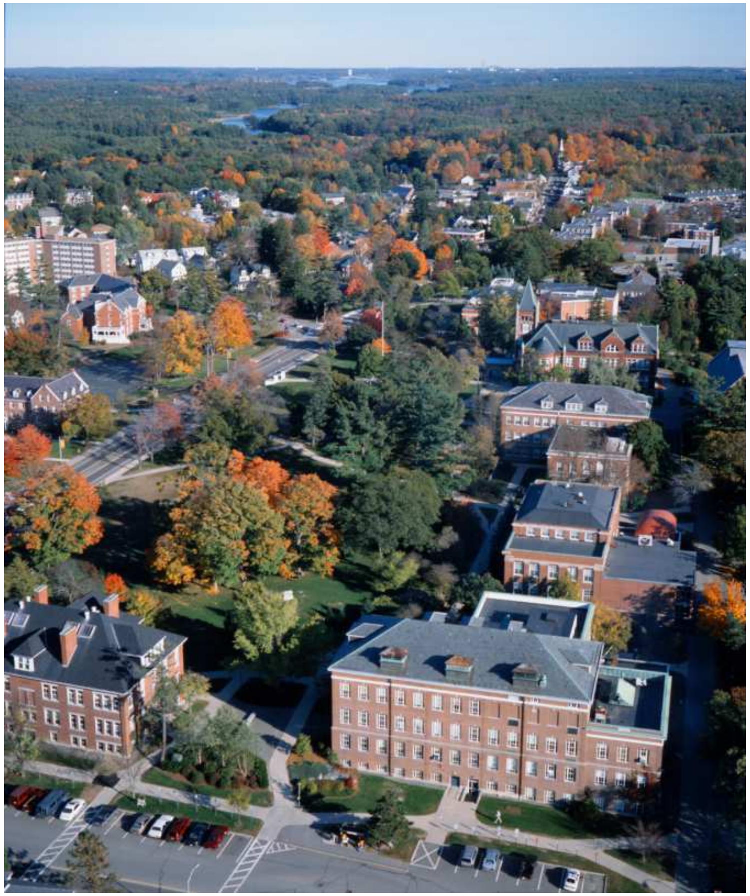
Summary

The OH-wOW framework is a:

- Fast,
- Simple,
- General,
- Online,
- Approximate,
- Way of Handling Open Worlds.

The University of New Hampshire

Tell your students to apply to grad school in CS at UNH!



- friendly faculty
- funding
- individual attention
- beautiful campus
- low cost of living
- easy access to Boston, White Mountains
- strong in AI, infoviz, networking, systems, bioinformatics