Planning over MDPs through Probabilistic HTNs

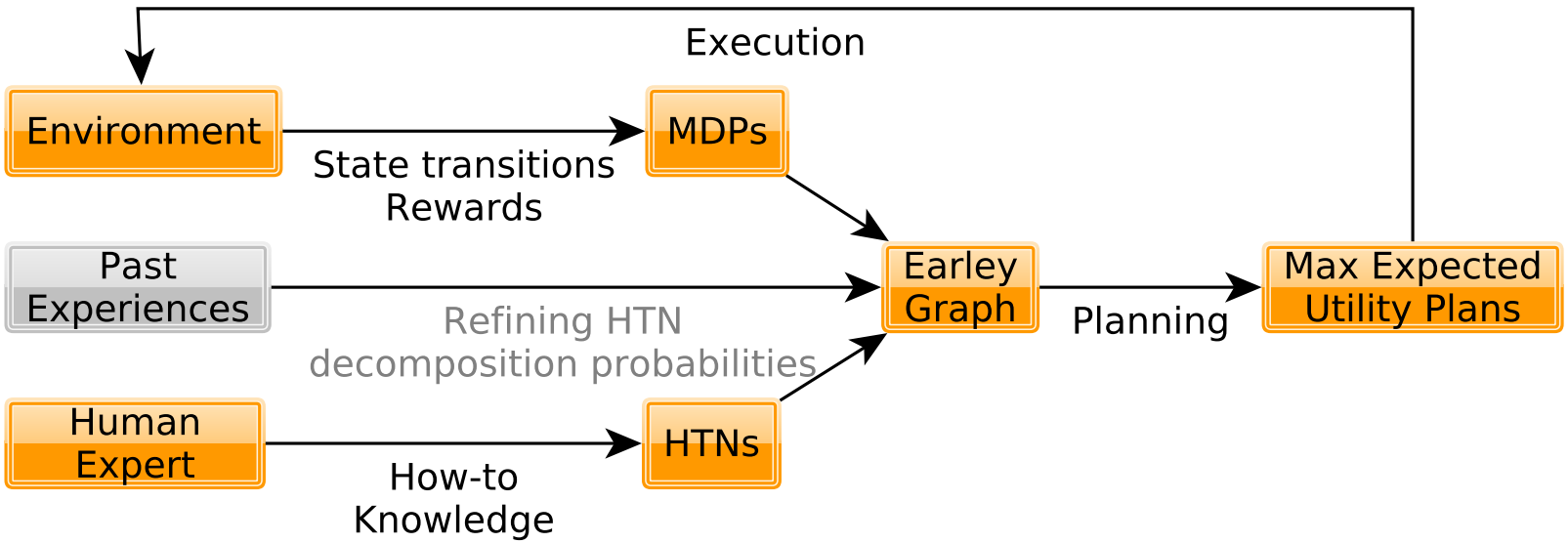
Yuqing Tang[1] Felipe Meneguzzi[3] Katia Sycara[3]
Simon Parsons[1,2]

[1]Graduate School and University Center
[2]Brooklyn College
City University of New York
ytang@cs.gc.cuny.edu
parsons@sci.brooklyn.cuny.edu

[3]Robotics Institute
Carnegie Mellon University
meneguzz,katia@cs.cmu.edu

August 8th, 2011

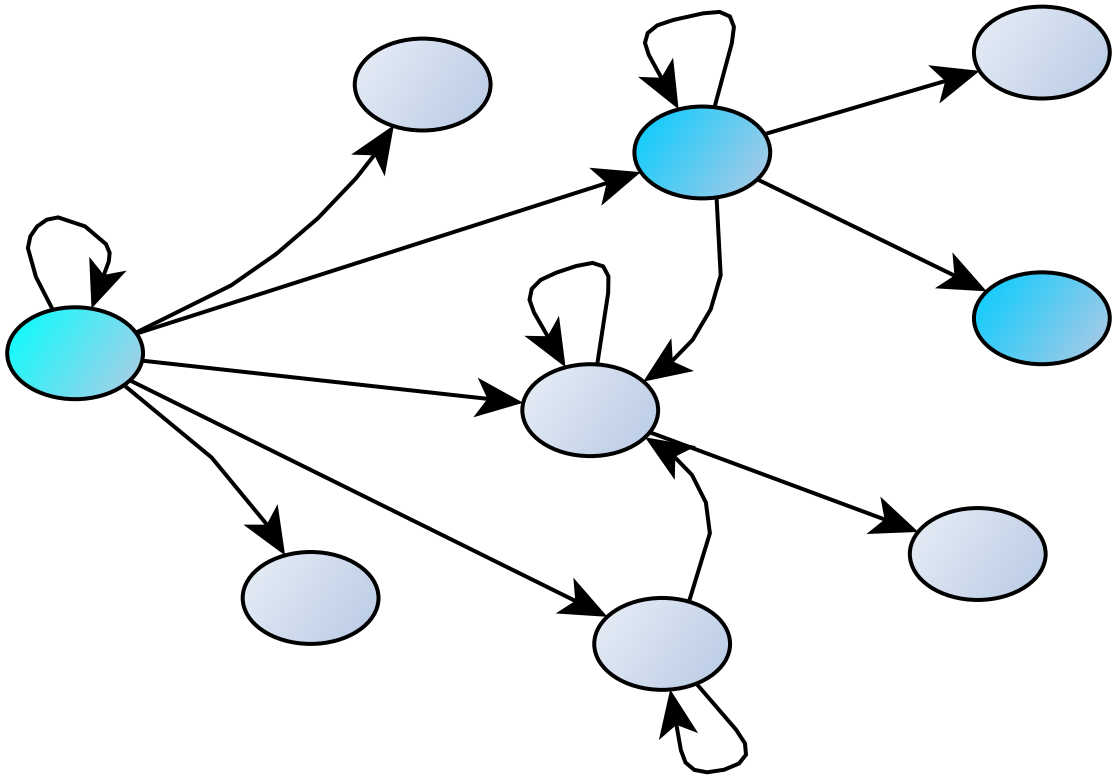
Motivations



- Planning when
  - The environment is non-deterministic — Markov Decision Processes (MDPs)
  - The applicability of the how-to knowledge on tasks is uncertain — probabilistic hierarchical task networks (HTNs)
- Earley graphs (borrowed from natural language processing [Stolcke, 1995]) to integrate probabilistic HTNs and MDPs

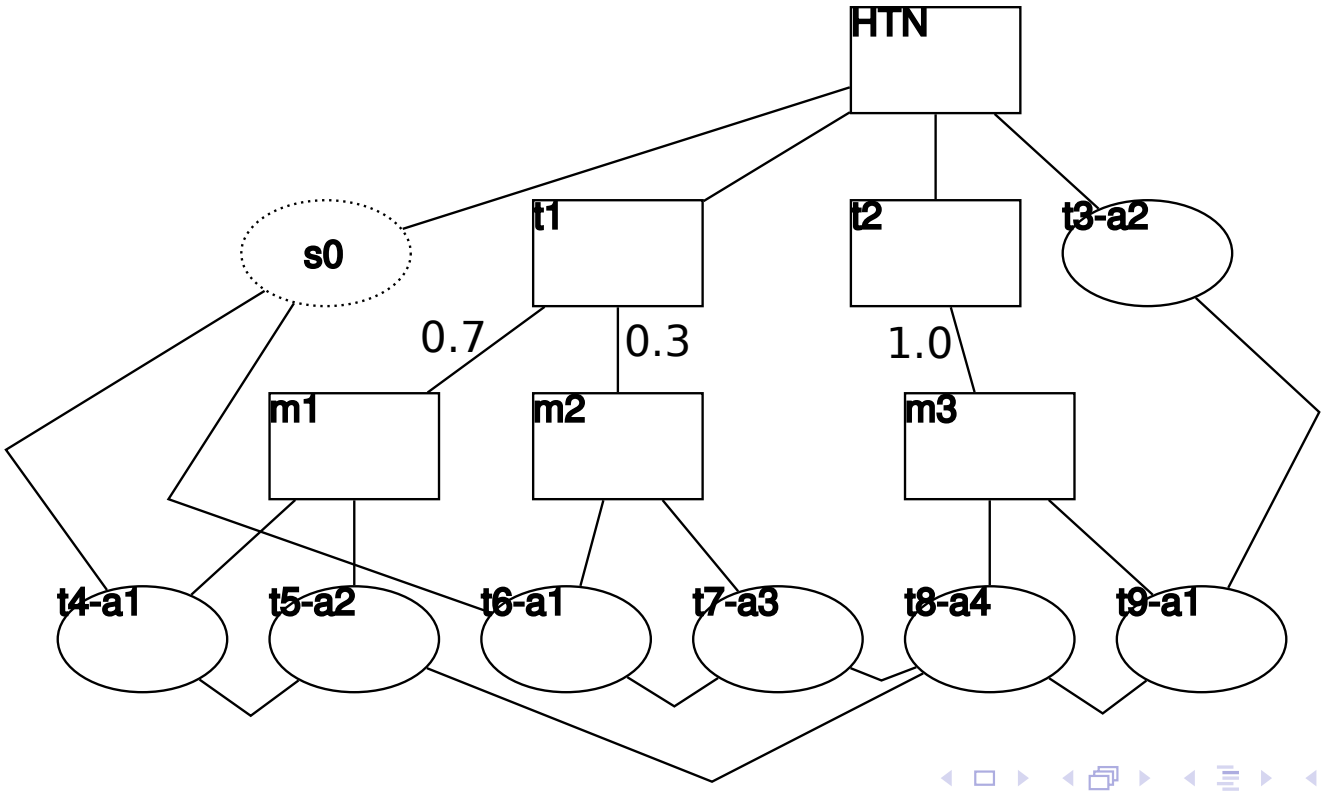
Markov Decision Processes

- Probabilistic state transitions: $Pr : S \times S \times A \rightarrow [0, 1]$
  - Markov Property: The resulting state is independent of previous history
- Rewards: $R : S \rightarrow [0, 1]$



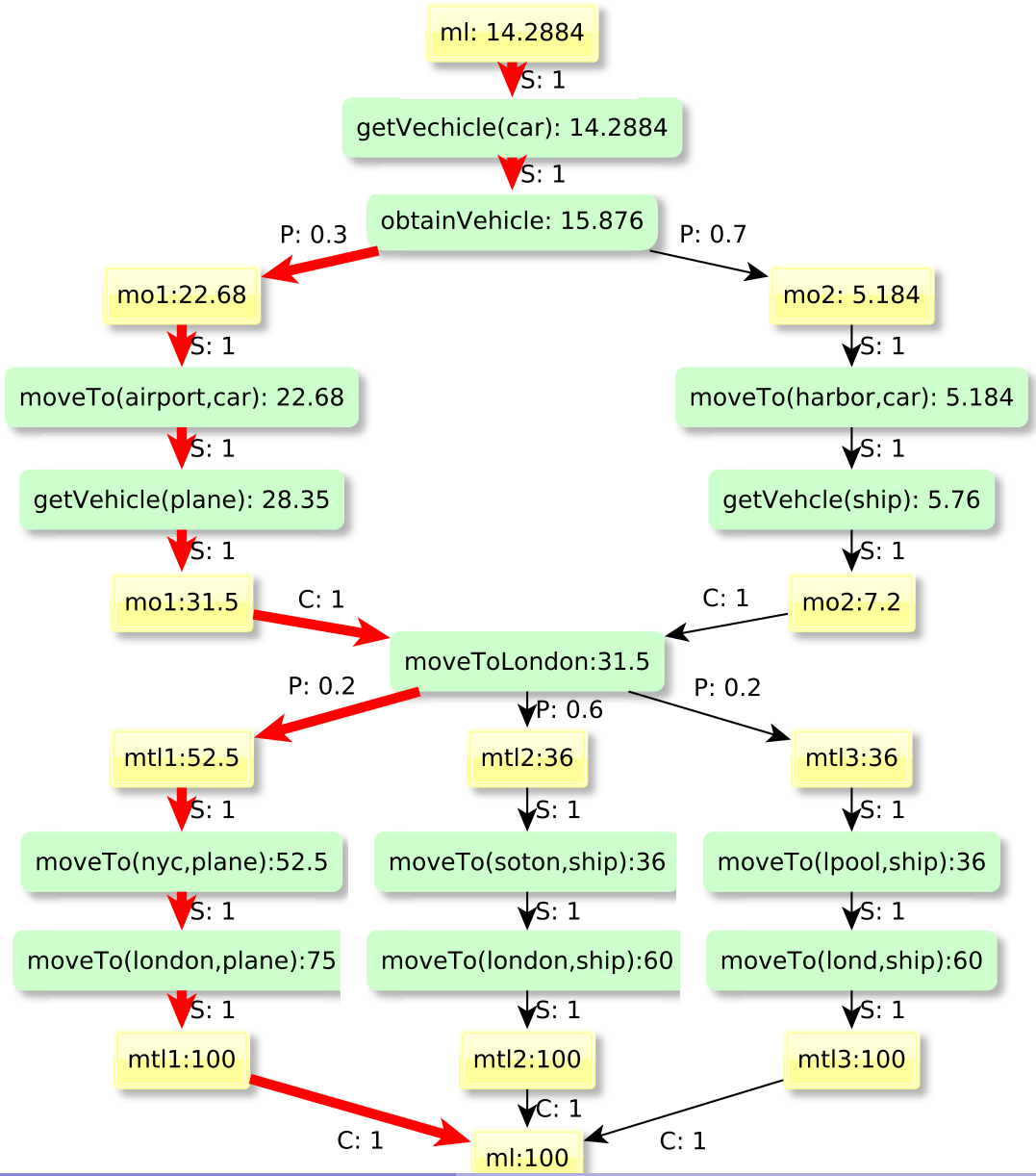
Probabilistic Hierarchical Task Networks (HTNs)

- Human experts abstract how-to-act knowledge into HTNs
- HTNs
  - Primitive tasks: Directly executable tasks (actions)
  - Non-primitive tasks: Symbolic abstraction about what to do
  - Methods: Recipes about how to decompose tasks into smaller tasks
    - ★ Decomposition probability: $[0, 1]$
  - Task networks: A set of tasks that are (partially) ordered or constrained



Probabilistic Earley Graph for HTNs

- Prediction links: Non-primitive task to method decomposition
- Scanning links: Primitive task executions
- Completing links: Completing methods preceding to next tasks
- #Nodes: $\Sigma_{m \in \mathcal{M}} |network(m)|$
- #Edges: $O\left(\Sigma_{m \in \mathcal{M}} |network(m)| + 2\Sigma_{t \in NT} \left(|Appearances(t)| \cdot |Methods(t)| \right)\right)$

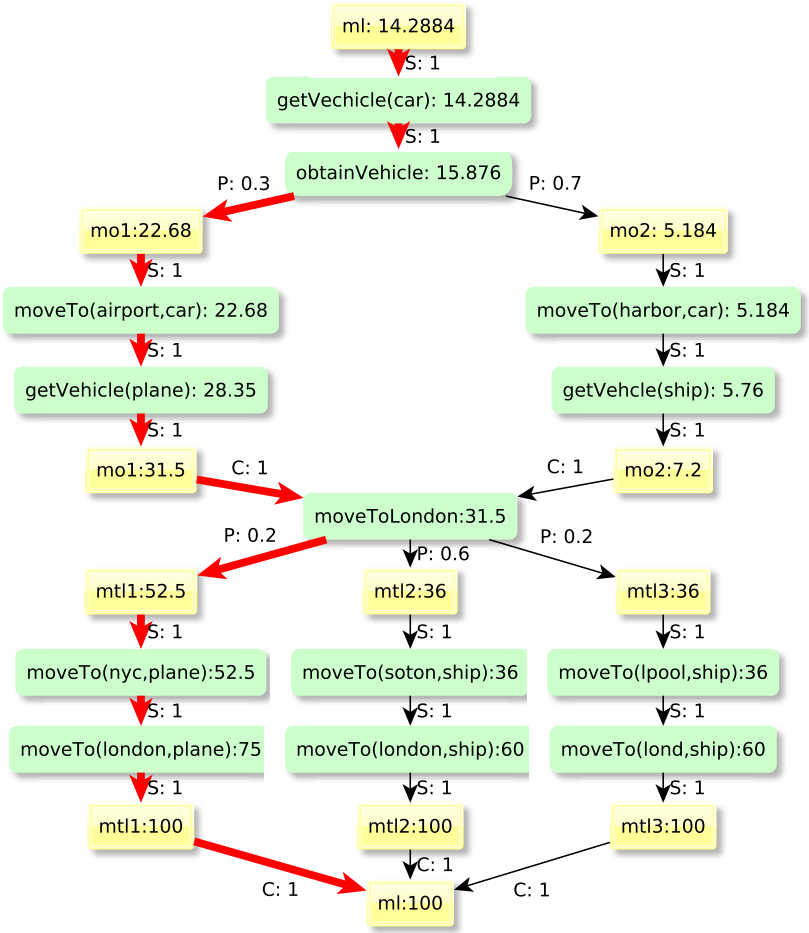


Computing Probabilities and Rewards with Earley Graph

The probability of a decomposition-execution path de

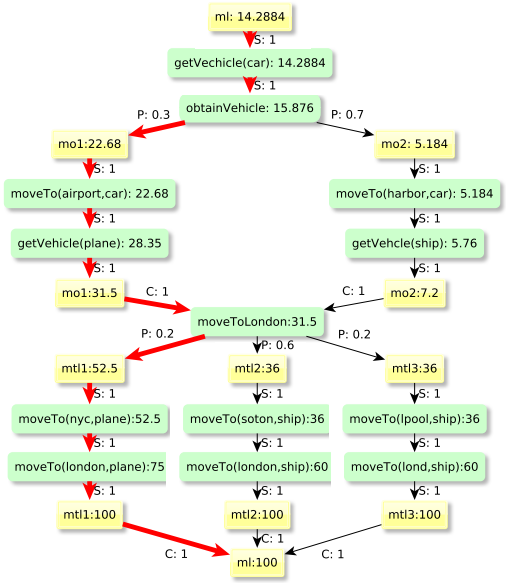
$$Pr(de) = \prod_{EN_i \in de} Pr(EN_{i+1}|EN_i) \cdot \prod_{s_j, a_j, s_{j+1} \in de} Pr(s_{j+1}|s_j, a_j)$$

- Compute probabilities of the decomposition-execution paths
- Compute expected rewards of decomposition-execution paths
- Learn decomposition probabilities from past experiences



Solution Concepts with Earley Graph

- Simple solution: A decomposition-execution path with maximum expected reward
- Generalized solution:
  - Given a sequence of encountered states and executed actions, compute the probability for each predictive nodes that the system can be in
  - Associate with each prediction node (non-primitive task) a table of rewards $\{\langle s, \text{method}, \text{method-parameters}, \text{expected reward}\rangle\}$ to select the method that can lead to MEU
  - Associate with each scanning node (primitive task) a table of $\{\langle s, \text{action-parameters}, \text{expected reward}\rangle\}$ to select the action parameters that can lead to MEU



Summary and Future Directions

- Earley graphs (borrowed from natural language processing [Stolcke, 1995]) to integrate probabilistic HTNs and MDPs
  - Compute probabilities and rewards of the decomposition-execution paths
  - MEU planning
- Future directions
  - Experiment with the generalized solution concepts
  - Learn decomposition probabilities
  - Bridge to plan recognition systems
  - Bridge to multiagent systems
    - ★ Agents observe other agents' behaviors and run plan recognition algorithms to obtain probabilistic HTNs
    - ★ Agents communicate to incrementally share the how-to knowledge (represented in HTNs) with a potential to converge into a same set of task networks
    - ★ Agents adapt to the environment and revise their task decomposition probabilities
    - ★ Agents converge to a set of cooperative behaviors regulated by the HTNs authorized by individual agents in an uncertain environment

Acknowledgments

Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

Andreas Stolcke.

An efficient probabilistic context-free parsing algorithm that computes prefix probabilities.

Computational Linguistics, 21(2):165–201, 1995.