Supplementary Material Learning for Decentralized Control of Multiagent Systems in Large Partially Observable Stochastic Environments

0.1 The Computation of Forward and Backward Variables (α, β)

The forward and backward variables (α, β) $\alpha_\tau^{n,k}(i) = p\big(q_{n,\tau}^k = i | m_{n,0:\tau}^k, o_{n,1:\tau}^k, \Theta_n\big)$ and $\beta_{t,\tau}^{n,k}(i) = \frac{p(m_{n,\tau+1:t}^k | q_{n,\tau}^k = i, o_{n,\tau+1:t}^k, \Theta_n)}{\prod_{\tau'=\tau}^t p(m_\tau^k | h_{n,\tau'}^k, \Theta_n)}$, $\forall\, n, k, t, \tau$ defined in section (MacDec-POMDP Policy Learning by Expectation Maximization) are similar to the forward-backward messages in the Baum-Welch algorithm for hidden Markov models [1]. These variables are computed recursively by each agent using (1)-(3).

$$\alpha_\tau^{n,k}(i) = \begin{cases} \frac{\mu(q_{n,0}^k = i)\lambda(q_{n,0}^k = i, m_{n,0}^k)}{p(m_{n,0}^k | h_{n,0}^k, \Theta_n)} & \text{if } \tau = 0 \\ \frac{\sum_{j=1}^{|Q_n|} \alpha(q_{i,\tau-1}^k(j)\delta(q_{i,\tau-1}^k = j, m_{n,\tau-1}^k, q_{n,\tau}^k, q_{n,\tau}^k = i)\lambda(q_{n,\tau}^k = i, m_{n,\tau}^k)}{p(m_{n,\tau}^k | h_{n,\tau}^k, \Theta_n)} & \text{if } \tau > 0 \end{cases} \tag{1}$$

$$\beta_{t,\tau}^{n,k}(j) = \begin{cases} \frac{1}{p(m_{n,0}^k | h_{n,0}^k, \Theta_n)} & \text{if } \tau = 0 \\ \frac{\sum_{n=1}^{|Q_n|} \delta(q_{n,\tau}^k = i, m_{n,\tau-1}^k, o_{n,\tau}^k, q_{n,\tau+1}^k = j)\lambda(q_{n,\tau+1}^k = j, m_{n,\tau+1}^k)\beta_{n,t,\tau+1}^k(j)}{p(m_{n,\tau}^k | h_{n,\tau}^k, \Theta_n)} & \text{if } \tau > 0 \end{cases} \tag{2}$$

$$p(m_{n,\tau}^k | h_{n,\tau}, \Theta_n) = \begin{cases} \sum_{j=1}^{|Q_n|} \mu(q_{n,0}^k = i)\lambda(q_{n,0}^k = i, m_{n,0}^k) & \text{if } \tau = 0 \\ \sum_{i,j=1}^{|Q_n|} \alpha(q_{n,\tau-1}^k = i)\delta(q_{n,\tau-1}^k = i, m_{n,\tau-1}^k, o_{n,\tau}^k, q_{n,\tau}^k = j)\lambda(q_{n,\tau}^k = j, m_{n,\tau}^k) & \text{if } \tau > 0 \end{cases} \tag{3}$$

0.2 Search and Rescue Domain

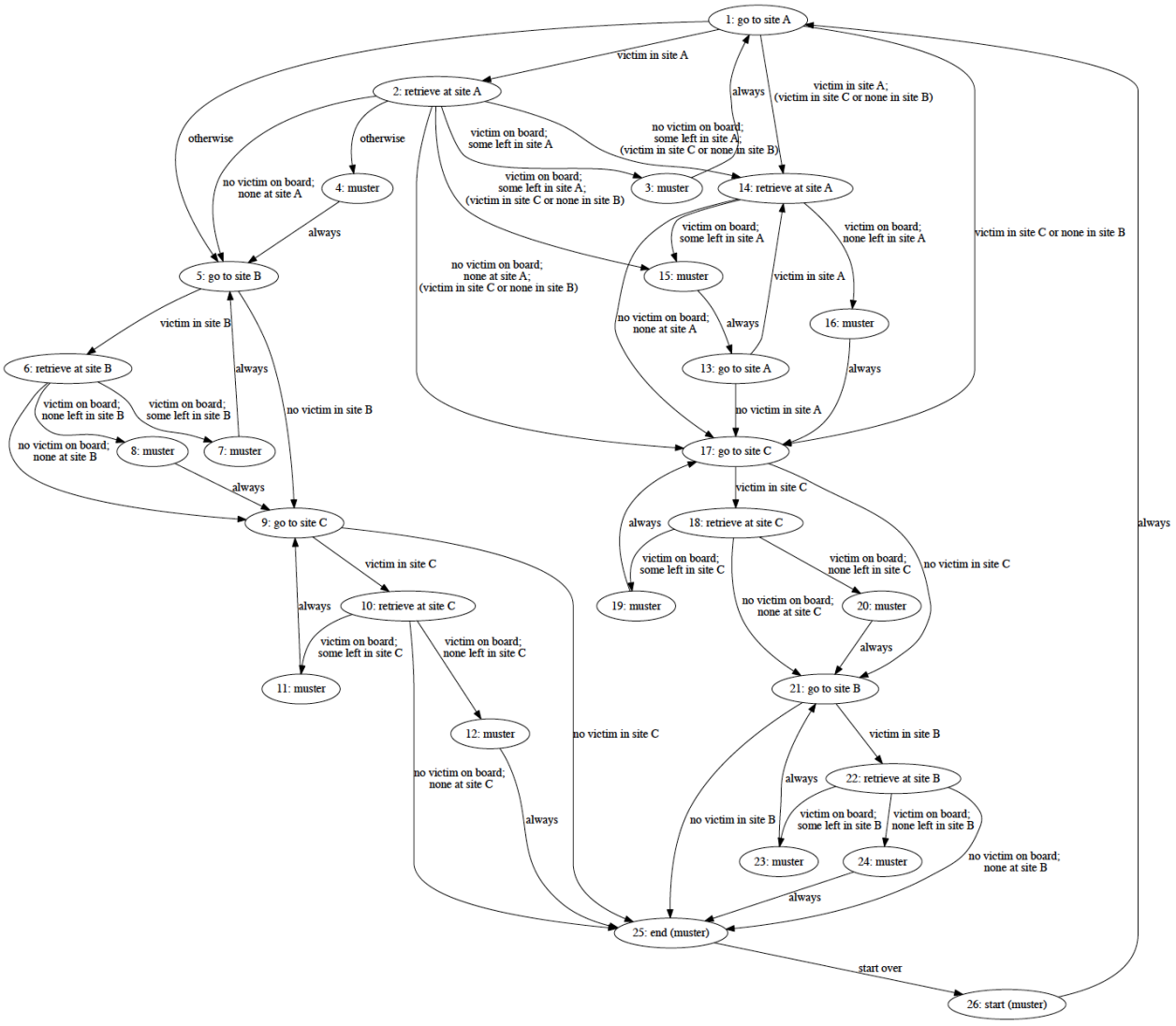

Figure 1: A heuristic policy finite state machine for UGV constructed by domain experts.

References

[1] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.