2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Stochastic Recruitment: Controlling State Distribution among Swarms of Hybrid Agents

Lael Odhner and Harry Asada

Abstract— **This paper introduces a control architecture for centrally controlling the ensemble behavior of many identical agents. A swarm of robots or other agents performing a variety of tasks is often modeled as a collection of hybrid-state agents, whose discrete switching behaviors are controlled by finite state machines. The number of agents in the swarm in a particular discrete state is a function of the rate at which agents transition between state. These state transitions are often modeled as stochastic interactions with the environment. We show that effective control over the distribution of agents in each discrete state can be achieved by designing the agents to transition between tasks randomly, according to a centrally determined state transition probability graph. The centrally-determined policy varies with time and with feedback information by re-broadcasting the probability graph to all agents. Feedback policies will be presented for the case in which the central controller has limited or no knowledge of the states of each agent.**

I. Introduction

There is an increasing interest in the robotics, biological engineering, and control communities in controlling collective behaviors of vast numbers of identical, independent units. In this paper we will focus on a specific class of problems in this area which we call recruitment problems, dealing with agents having a set of discrete states governing tasks or behaviors. Recruitment is the problem of centrally controlling the fraction of agents in each discrete state. The goal is to affect this ensemble distribution of agents in any particular discrete state, using an extremely low-bandwidth broadcast command and limited feedback information, as shown in Fig. 1. This architecture is inspired by the example of skeletal muscle, which is composed of many small independent sub-systems called motor units. The nervous system stimulates the motor units, which individually relax or contract to produce a summed output force based on an individual response threshold [1]. This idea of recruitment is particularly applicable to problems having a hybrid state description, that is, a discrete finite state machine which controls the switching behavior of a robot or agent between several different state evolution equations. The aforementioned problem of motor recruitment could be discretized into many hybrid state machines whose state transitions are triggered by nervous activation at each motor unit. Similarly, the collective behaviors of insects, or even cellular regulatory mechanisms can be treated as systems composed of many hybrid-state units [2] [3].

The authors are with the Mechanical Engineering Department at the Massachusetts Institute of Technology, Cambridge, MA, USA, 02139 {lael, asada}@mit.edu

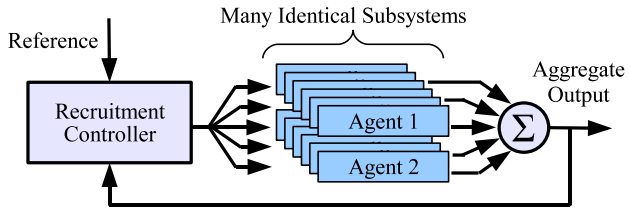

Fig. 1. A block diagram representation of the recruitment control problem. One central controller must send the same signal to a large swarm of identical agents, in order to control the distribution of states among the swarm.

Many researchers have described robotic swarms as a collection of hybrid-state agents. Each different behavior performed by the agents in the swarm has its own discrete state, and the interactions with the environment or other robots that trigger transitions between behaviors are modeled using a Markov probabilistic model. In the example shown in Fig. 2, a robot carrying out a task to search for and retrieve an object will transition between its "search" and "carry" states based on the random event of finding the object it seeks in the environment. Over time, these stochastic transitions often reach a steady state equilibrium distribution, similar to a chemical reaction or other kinetic process. These models have been successful in predicting the emergent behavior of swarms [4] [5] [6], and have also been used to generate rules for each agent to follow in transitioning between tasks.

This framework can also be applied to systems that have been intentionally hybridized, such as active material actuators that have been intentionally broken up into very many on/off units similar to motor units in biological muscle. The motivation for imposing this architecture is usually that designing a two-state regulator is simpler or more robust than continuously varying the response of an active material[7] [8] [9]. In swarm robot systems, it has been noted that intentional pseudo-random transitions can generate useful behaviors [10] [11]. Unlike most hybrid state systems, in which probabilistic state transitions result from interactions with the environment, the authors have been focusing on systems in which each agent is designed to transition in a pseudo-random manner between all states. The Markov state transition graph associated with these transitions is chosen by a central controller and broadcast to all of the agents. This transition graph constitutes the control input for the system, and can vary over time according to a control policy. This kind of stochastic recruitment strategy requires very little

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

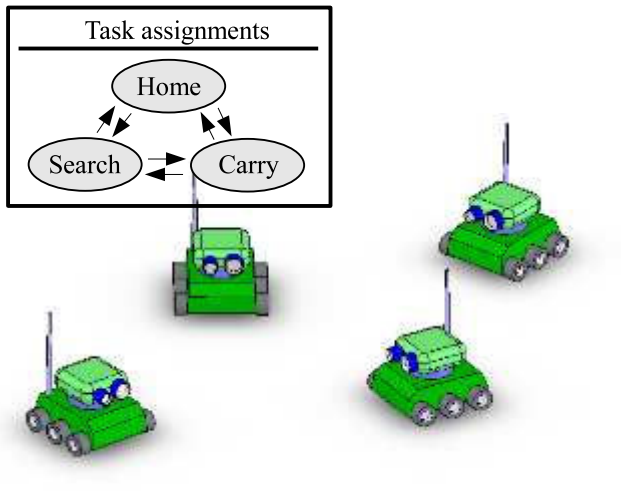

Fig. 2. Many models of swarm behavior include a finite state machine description of each agent in the swarm. The transitions between states are governed by random interactions between agents and the environment or agents and other agents.

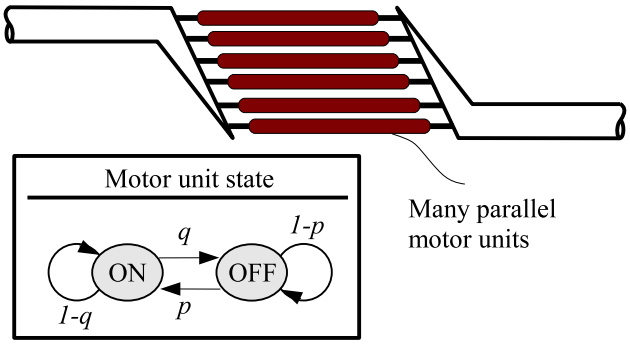

Fig. 3. The motor recruitment problem in an artificial muscle can be posed as a set of motor units having two states, ON and OFF, in which different levels of force are produced. The net force produced is the sum of each individual unit.

bandwidth, and renders the agents in the swarm anonymous and interchangeable.

The authors have previously analyzed problems of recruitment in which the distribution of agents in each state is fully known. In this case, both linear and time-optimal feedback policies can be formulated to drive the state distribution towards a desired reference [12] [13] [14]. In this paper we will introduce control policies for the case when the central controller has limited knowledge or no knowledge of the exact number of agents in each state. In the case where the central controller has no knowledge of the agents' state, a class of control policies based on simple kinetic equilibrium predictions allows the central controller to specify any distribution and achieve it within some variance. Time-optimal control laws are also derived for a system in which minimal feedback knowledge is available. This analysis is particularly useful because it ties together the feed-forward and feedback policies, and also provides a good bounding case on the behavior of feedback policies.

II. The Dynamics of Randomized Recruitment

Consider the simplest recruitment problem, a system made up of N agents having just two states, ON and OFF. The control task is to recruit a specific number of agents, N^{ref}, into the ON state. These states could represent two different sensor modalities or behaviors exhibited by a robot. The authors have been using two-state agents of this kind to control shape memory alloy actuators made up of many binary units, capable of producing varied force and displacement as a function of the number of ON agents, N_t^{on}. Figure 3 shows a schematic of this actuator architecture and the Markov graph governing the transition between discrete states. To keep track of the state evolution of the system, we will introduce a discrete distribution variable, $\mathbf{x}$, describing the probability with which each agent is in either state,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} P(ON) \\ P(OFF) \end{bmatrix} \tag{1}$$

In previous publications, we have considered the case where control policies have explicit knowledge of N_t^{on}[12]. Here that assumption will be relaxed so that the distribution of N_t^{on} is predicted conditioned on $\mathbf{x}_t$ using a binomial distribution,

$$P(N_t^{on} = k|\mathbf{x}_t) = \begin{pmatrix} N \\ k \end{pmatrix} x_{1,t}^k (1 - x_{1,t})^{N-k} \tag{2}$$

Equation (2) will actually describe the number of agents in a selected state for a system with more than two states. The distribution can be viewed as a Bernoulli process summing a random variable equal to 1 if the agent is in the selected state, and 0 otherwise. As N becomes large, the central limit theorem will guarantee that N_t^{on} will approach its expected value,

$$E(N_t^{on}|\mathbf{x}_t) = N x_{1,t} \tag{3}$$

The advantage of using $\mathbf{x}_t$ to keep track of the recruited units rather than directly using N_t^{on} is the simplicity of the state evolution model afforded by the probability distribution. The state transition graph of each agent is parametrized as shown in Figure 3, using variables p and q to represent the probabilities of transitioning from OFF to ON and ON to OFF, respectively. These variables make up the control input broadcast by the central controller. For the moment they will be determined by a constant policy, but later it will be clear that there is often benefit to be gained by varying p and q over time. The evolution of $\mathbf{x}$ can be written as a matrix $\mathbf{M}$,

$$\mathbf{x}_t = \mathbf{M}^t \mathbf{x}_0 = \begin{bmatrix} 1-q & p \\ q & 1-p \end{bmatrix}^t \mathbf{x}_0 \tag{4}$$

The eigenvalue-eigenvector decomposition of $\mathbf{M}$ can be used to produce a simpler representation of (4),

$$M = \begin{bmatrix} \frac{p}{p+q} & 1 \\ \frac{q}{p+q} & -1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & 1-p-q \end{bmatrix} \begin{bmatrix} \frac{p}{p+q} & 1 \\ \frac{q}{p+q} & -1 \end{bmatrix} \quad (5)$$

This decomposition can then be used to rewrite (4),

$$\mathbf{x}_t = \mathbf{x}_{ss}\lambda_1^t + \mathbf{x}_{trans}\lambda_2^t \quad (6)$$

As is the case with any conservative Markov model, the largest eigenvalue of $\mathbf{M}$, λ_1, is equal to 1. The steady-state distribution $\mathbf{x}_{ss}$ is found in the corresponding eigenvector,

$$\mathbf{x}_{ss} = \begin{bmatrix} \frac{p}{p+q} \\ \frac{q}{p+q} \end{bmatrix} \quad (7)$$

Some physical meaning can be gleaned from (7). It states that the fraction of ON agents at steady state is equal to the probabilistic rate at which agents transition from OFF to ON, normalized by the sum of all transition rates between states. The second eigenvalue of $\mathbf{M}$, λ_2, varies between -1 and 1 and is equal to $1-p-q$, and its eigenvector, $\mathbf{x}_{trans}$, makes up the transient portion of the response in Equation (6). If the system has more than two states, the second-largest eigenvalue of $\mathbf{M}$ will dominate the settling time.

III. A NO-KNOWLEDGE CONTROL POLICY

If the central controller has no knowledge of the number of agents in each state, then the control policy must produce feed-forward dynamics that move the state distribution toward the desired goal. Equation (7) demonstrated that the recruitment dynamics have a stable steady-state component in the feed-forward response, so p and q could be chosen so that some desired number of cells N^{ref} is expected according to (3),

$$\frac{p}{p+q} = \frac{N^{ref}}{N} \quad (8)$$

Many policies satisfy this constraint. For example, setting $p = 0.1$ and $q = 0.1$ will drive the agents to a 50% likelihood of being in either state according to (7). So will setting $p = q = 0.3$. In order to distinguish between these cases, a scaling analysis can be used to weigh the performance tradeoffs between these policies.

A. Convergence Rate and Steady State Distribution are Independent.

The control policy parameters p and q can be rewritten as βp_0 and βq_0, where $p_0 + q_0 = 1$ and β is a scaling factor that varies between 0 and $\max(1/p_0, 1/q_0)$. When the transition probabilities are scaled in this way, the steady-state distribution $\mathbf{x}_{ss}$ from (7) is independent of β:

$$\mathbf{x}_{ss} = \begin{bmatrix} \frac{\beta p_0}{\beta(p_0+q_0)} \\ \frac{\beta q_0}{\beta(p_0+q_0)} \end{bmatrix} = \begin{bmatrix} \frac{p_0}{p_0+q_0} \\ \frac{q_0}{p_0+q_0} \end{bmatrix} \quad (9)$$

However, the smaller eigenvalue governing the rate of convergence still depends on β:

$$\lambda_2 = 1 - \beta(p_0 + q_0) = 1 - \beta \quad (10)$$

This means that β is a free parameter with which the convergence time can be arbitrarily varied while still satisfying the condition imposed in (8). In the special case $\beta = 1$, the second eigenvalue $\lambda_2 = 0$. In this case, $\mathbf{x}$ converges to $\mathbf{x}_{ss}$ after only one round of random state transitions. Figure 4 shows $\mathbf{x}_t$ converging to the same steady-state behavior from the same initial conditions, for several values of β.

B. Accuracy Varies Only as $\boldsymbol{x}_{ss}$ and N.

Specifying $\mathbf{x}_{ss}$ does not guarantee that the number of recruited cells N_{ss}^{on} will converge to the desired number. The accuracy of the control system once it has reached steady state can be described by the variance of N_{ss}^{on}, normalized by the number of agents N,

$$Var\left\{\frac{N_{ss}^{on}}{N}|\mathbf{x}_{ss}\right\} = \frac{x_{1,ss}(1-x_{1,ss})}{N} = \frac{pq}{(p+q)^2 N} \quad (11)$$

In order to extend this expression into the many-state case, all that is needed is the expression for $x_{1,ss}$, an element of the stable eigenvector of $\mathbf{M}$. For the two-state case, the β scaling argument from (9) and (10) can be applied to the variance calculation. The numerator and denominator of (11) both vary by a factor of β^2, so the variance is independent of the rate at which the actuator converges to its steady state probability distribution,

$$Var\left\{\frac{N_{ss}^{on}}{N}|\mathbf{x}_{ss}\right\} = \frac{\beta^2 p_0 q_0}{\beta^2(p_0+q_0)^2 N} = \frac{p_0 q_0}{(p_0+q_0)^2 N} \quad (12)$$

This is an important observation; it means that nothing is to be gained by taking "baby steps", that is, selecting very small values of p and q in hopes of improving the accuracy of recruitment in exchange for a slower rate of response. It also means that, for any desired number of recruited agents, the only way to improve the accuracy of this control system is to increase the number of agents, N.

C. The Number of Transitions per Unit Time

In a physical system, there is often a significant energy cost associated with switching agents from one behavior to another. For example, a mobile robot switching between patrolling two different areas will expend energy in driving from place to place. A shape memory alloy actuator has significant latent heat associated with the phase transition used for actuation, so spurious phase transitions are costly. As a consequence, it may be useful to consider the expected number of state transitions per unit time when formulating a control policy. The expected number of transitions can be calculated conditioned on $\mathbf{x}_t$, p and q,

$$E(N_t^{trans}|\mathbf{x}_t, p, q) = N(qx_{1,t} + px_{2,t}) \quad (13)$$

In the steady state (8) can be substituted in, so that (13) is a function of N, p and q,

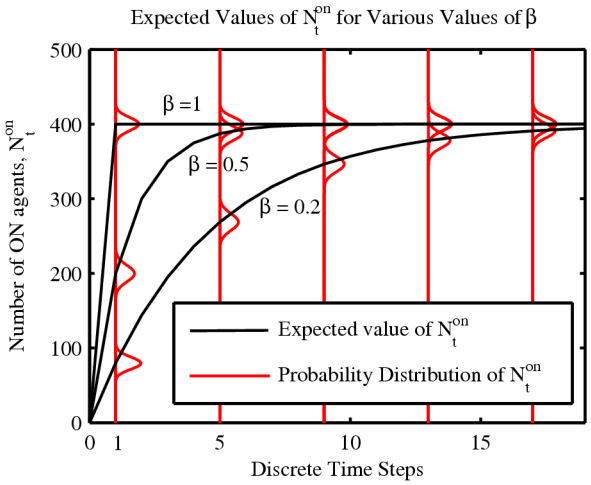

Fig. 4. The expected value of N_t^{on} and probability distribution of N_t^{on} at several points in time are shown for $N = 500$, $p_0 = 0.8$, $q_0 = 0.2$, and β = 0.2, 0.5 and 1. This plot illustrates the fact that the variance of N_t^{on} is independent of the rate of convergence. All three cases approach the same probability distribution in N_t^{on}.

TABLE I

SCALING OF PERFORMANCE MEASURES VERSUS β

Performance Measure	Dependency on β	Goal
$E(N_{ss}^{on})$	none	N^{ref}
$Var(N_{ss}^{on}/N)$	none	0
Convergence Rate λ_2	$1-\beta$	0
$E(N_{ss}^{trans})$	$\beta(2Np_0q_0)$	0

$$E(N_{ss}^{trans}|p,q) = \frac{2Npq}{p+q} \tag{14}$$

Using the scaling argument again, (14) can be rewritten in terms of βp_0 and βq_0. This implies that an increase in β implies more expected transitions per unit time in the steady state,

$$E(N_{ss}^{trans}|\beta p_0, \beta q_0) = \frac{\beta^2 2Np_0q_0}{\beta(p_0+q_0)} = \beta(2Np_0q_0) \tag{15}$$

The value of β minimizing the number of expected transitions is, naturally, 0, corresponding to the control policy that allows no random transitions between states.

D. Summary of Trade-Offs

The results of the scaling analysis, shown in Table I, show that there is a conflict between the rate at which the system converges, and the expected number of state transitions that occur at steady state. Choosing $\beta = 1$ will yield the fastest convergence of $\mathbf{x}_t$ to $\mathbf{x}_{ss}$, but it will also incur a substantial number of state transitions per unit time. Knowledge of these trade-offs is important because it can enable the selection of higher-performing, time-varying policies. For example, if it is acceptable to have a small constant error in the state distribution, a compromise policy might set p and q to satisfy (8) and $\beta = 1$, then cease all transition after one round of stochastic state transitions, once $\mathbf{x} = \mathbf{x}_{ss}$. The variance of the error in this policy would be equal to the steady state variance for the constant policy, while eliminating many state transitions.

IV. A MINIMAL KNOWLEDGE FEEDBACK POLICY

The analysis above provides some insight into the performance capabilities and performance limitations of the open-loop system. Although the expected distribution of the open-loop system converges, there is undesirable uncertainty in the response. The obvious means for improving on the performance of the open-loop response is the incorporation of feedback. Here we will analyze the simplest possible feedback policy, in which the central controller knows only when the distribution of agents is close enough or equal to N^{ref}. This knowledge is represented as a Boolean variable y_t,

$$y_t = \begin{cases} true, & N_t^{on} = N^{ref} \\ false, & N_t^{on} \neq N^{ref} \end{cases} \tag{16}$$

This extremely limited amount of information will produce policies of limited practical application, but this solution is conceptually important for two reasons. First, this is in some sense the worst-case situation for which probability one convergence is assured. As such, it provides a useful bounding case on the more general problem of limited information control. Second, the optimal policy based on this limited feedback is closely related to the open-loop control policy, and thus provides a link between the behavior of limited-knowledge and no-knowledge policies.

A. Probability One Convergence

One of the simplest requirements for a feedback control policy is that it converge in probability, that is, that the probability of reaching the desired state monotonically approaches one with time. Given only y_t, this can be achieved by choosing policies which cease all state transitions (command $p = q = 0$) upon reaching the desired distribution. A brief proof follows.

Proof of P=1 Convergence: For any control policy in which p and q are greater than zero by some non-infinitesimal amount, the agents will be ON or OFF with some probability distribution, $\mathbf{x}_t$ after at least one round of stochastic state transitions. According to the probability distribution in (2), the system has a non-infinitesimal, non-zero probability of achieving any value of N_t^{on}. The greatest lower bound for all these probabilities will be called ρ. If the control policy is designed to stop, that is, set $p = q = 0$ when y_t indicates that the current state is the target state, the probability of leaving the target state is zero. The net probability of being in the target state after k time steps is consequently greater than or equal to $1-(1-\rho)^k$. As k becomes very large, this quantity approaches 1 exponentially (end of proof).

Because this class of policies creates a de facto absorbing target state, this problem can be approached as a stochastic shortest path (SSP) problem in the dynamic programming

framework. We will derive a control policy that minimizes the expected time to converge. A non-discounted, positive cost-per-stage function $g(y_t)$ can be written as a random variable keeping track of the time spent not in the target distribution,

$$g(y_t) = \begin{cases} 0, & y_t = true \\ 1, & y_t = false \end{cases} \tag{17}$$

The total cost function is the convergence time of the system written as a summation of $g(y_t)$,

$$J(X_0) = E\left\{ \sum_{t=0}^{\infty} g(y_t) \middle| \mathbf{x}_0 \right\} \tag{18}$$

Because the cost-per-stage $g(y_t)$ is bounded and all admissible control policies converge with probability one, J is convergent. As in the no-knowledge case, nothing is known about the number of ON agents until the target distribution is achieved, so a policy with constant p and q must be pursued until y_t is true. The optimal policy to pursue is not hard to imagine; Essentially, it is to set p and q to satisfy (8) and $\beta = 1$ until y_t is true.

Proposition: The control policy which minimizes (18) for any initial value of $\mathbf{x}_0$ is the policy which sets $x_{1,ss} = N^{ref}/N$ and $\lambda_2 = 0$ unless y_t indicates that the target state has been reached, in which case $p = q = 0$.

Proof: This policy is a minimum of J because at $x_{1,ss} = N^{ref}/N$ and $\lambda_2 = 0$, the partial derivatives of J with respect to $x_{1,ss}$ and λ_2 are increasing functions equal to zero, and consequently there is a minimum of J at this policy. The cost function can be rewritten using Bellman's equation,

$$J(\mathbf{x}_t) = E\left\{ g(y_t) + J(\mathbf{x}_{t+1}) \middle| \mathbf{x}_t \right\} \tag{19}$$

Using (2) and (16), we can write the probability $H(\mathbf{x}_t)$ that y_t is true conditioned on x_t,

$$H(\mathbf{x}_t) = P(y_t|\mathbf{x}_t) = \begin{pmatrix} N \\ N^{ref} \end{pmatrix} x_{1,t}^{N^{ref}} (1 - x_{1,t})^{N - N^{ref}} \tag{20}$$

We can then rewrite (19) in terms of (20),

$$J(\mathbf{x}_t) = (1 - H(\mathbf{x}_t))(1 + J(\mathbf{x}_{t+1})) \tag{21}$$

The partial derivative of $J(x_{1,t})$ with respect to $x_{1,ss}$ can be found when $\lambda_2 = 0$ by differentiating (21),

$$\frac{\partial J(\mathbf{x}_t)}{\partial x_{1,ss}} = -\frac{\partial H(\mathbf{x}_t)}{\partial x_{1,ss}} J(\mathbf{x}_{t+1}) + (1 - H(\mathbf{x}_t)) \frac{\partial J(\mathbf{x}_{t+1})}{\partial x_{1,ss}} \tag{22}$$

The chain rule can then be used to evaluate $\partial H/\partial x_{1,ss}$,

$$\frac{\partial H}{\partial x_{1,ss}} = -\frac{\partial H}{\partial x_{1,t}} \frac{\partial x_{1,t}}{\partial x_{1,ss}} \tag{23}$$

The derivative of $H(\mathbf{x}_t)$ with respect to $x_{1,t}$ is found to be a product of $H(\mathbf{x}_t)$ and a second term,

$$\frac{\partial H}{\partial x_{1,t}} = \begin{pmatrix} N \\ N^{ref} \end{pmatrix} x_{1,t}^{N^{ref}} (1 - x_{1,t})^{N - N^{ref}} \left[\frac{N^{ref} - x_{1,t} N}{x_{1,t}(1 - x_{1,t})} \right] \tag{24}$$

The eigenvalue decomposition of $\mathbf{x}_t$ can be substituted into (24), holding λ_2 equal to 0,

$$\frac{\partial H}{\partial x_{1,t}} = H(\mathbf{x}_t) \left[\frac{N^{ref} - x_{1,ss} N}{x_{1,t}(1 - x_{1,t})} \right] \tag{25}$$

The derivative of $x_{1,t}$ with respect to $x_{1,ss}$ is 1. $H(\mathbf{x}_t)$ and $x_{1,t}(1 - x_{1,t})$ are both positive quantities, so $\partial H/\partial x_{1,ss}$ can be written as a positive quantity A_t multiplied by $N^{ref} - x_{1,ss}N$,

$$\frac{\partial H}{\partial x_{1,ss}} = A_t (N^{ref} - x_{1,ss} N) \tag{26}$$

Equation (22) can be evaluated using (26),

$$\frac{\partial J(\mathbf{x}_t)}{\partial x_{1,ss}} = (x_{1,ss} N - N^{ref}) A_t J(\mathbf{x}_{t+1}) + (1 - H(\mathbf{x}_t)) \frac{\partial J(\mathbf{x}_{t+1})}{\partial x_{1,ss}} \tag{27}$$

Each term of (27), when expanded, will contain a factor of $x_{1,ss}N - N^{ref}$. All of the other quantities in each term, values of A_t, J, and $1 - H(\mathbf{x}_t)$, are positive, so the sign of $x_{1,ss}N - N^{ref}$ will determine the sign of the entire expression. Consequently, $\partial J/\partial x_{1,ss}$ will be an increasing function of $x_{1,ss}$, equal to zero when $x_{1,ss} = N^{ref}/N$. A similar argument can be made for λ_2 when $x_{1,ss} = N^{ref}/N$. The partial derivative of $x_{1,t}$ with respect to λ_2 is equal to $t\lambda^{t-1}$. Using the chain rule, the $\partial H/\partial \lambda_2$ can be found when $x_{1,ss} = N^{ref}/N$,

$$\frac{\partial H}{\partial \lambda_2} = H(\mathbf{x}_t) \left[\frac{-t x_{trans} \lambda^{2t-1}}{x_{1,t}(1 - x_{1,t})} \right] \tag{28}$$

This expression can be written as the product of some positive quantity B_t and $-\lambda_2^{2t-1}$, much like (26),

$$\frac{\partial H}{\partial \lambda_2} = -B_t \lambda_2^{2t-1} \tag{29}$$

The derivative of J with respect to λ_2 can be written in a fashion similar to (27),

$$\frac{\partial J(\mathbf{x}_t)}{\partial \lambda_2} = \lambda^{2t-1} B_t J(\mathbf{x}_{t+1}) + (1 - H(\mathbf{x}_t)) \frac{\partial J(\mathbf{x}_{t+1})}{\partial \lambda_2} \tag{30}$$

Each term in this recursively defined series is an odd increasing function of λ. Because (27) and (30) are both increasing functions which are zero for $x_{1,ss} = N^{ref}/N$ and $\lambda = 0$, this policy minimizes the time it takes the swarm of agents to converge on the target distribution.

B. A Computational Example

Figure 5 shows a histogram of 10000 simulations of the time-optimal minimal feedback policy, for $N = 500$ and $N^{ref} = 200$, showing the typical convergence behavior. All simulations were started from the same initial conditions, with all agents in the OFF state. The optimal law, in which

$\beta = 1$, is compared to a more gradual law, having $\beta = 0.1$. In both cases, the likelihood of converging at any point in time should converge to a constant value, as x_t approaches x_{ss}. One would expect that the tail of both distributions should have an exponential shape, which the figure clearly demonstrates. The shifting of the distribution peak towards longer convergence time is due to the fact that the likelihood of reaching the desired output is very small until x_t is close to x_{ss}, which occurs quickly with the optimal policy and more slowly with the $\beta = 0.1$ feedback policy. The high variance of these convergence times, represented by the length of the exponential tails, is one of the major factors which can be mitigated by incorporating more information into the control policy. The expected value taken from the $\beta = 1$ distribution was 27.5 time intervals. The authors' previous work found that the expected convergence time under these same conditions with full knowledge of N_t^{on} was about 4.5 time intervals[1] [12]. A well-designed practical control policy should lie somewhere between these two extremes.

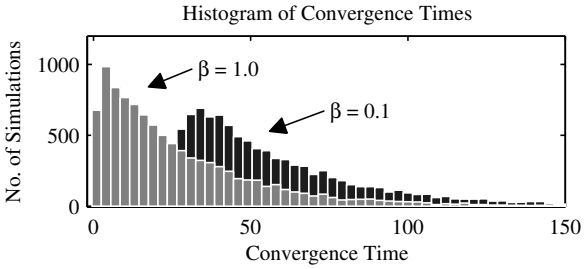

Fig. 5. A histogram showing the observed convergence time of the feedback policy for many trials

V. Conclusion

In this paper, we have shown that:

- It is possible to control the ensemble behavior of many finite-state agents by intentionally randomizing the behavior of each agent according to a centrally determined policy.
- If the central controller has no information about the state of the agents, control policies can be formulated which approximately achieve the desired distribution of states among agents.
- There is no benefit to taking "baby steps" toward the desired distribution; the one-shot policy which causes the individual probability distribution of each agent $\mathbf{x}_t$ to converge immediately provides the best accuracy achievable.
- If the central controller has minimal information about the state of the agents, policies can be formulated to cause the distribution to converge exponentially on the target distribution of states among agents.
- The policy which minimizes the expected convergence time of the system can be provably found for any desired distribution.

[1]These results were found using the value iteration algorithm, for a controller with measurements of the exact number of ON agents.

Although the analysis presented here is restricted to the two-state case, it can be extended to the many-state recruitment problem. This analysis is also restricted to very limited knowledge of the state of agents in the swarm. Past work on recruitment policies having full knowledge of N_t^{on} has yielded good results in finding numerically optimal control policies. Policies for systems in which the state of the agents is partially known, or in which the state is affected by random interactions with the environment, remain an interesting future directions.

VI. Acknowledgments

This material is based upon work supported by the National Science Foundation under grant number 0143242.

References

[1] E. Henneman, G. Somjen, and D.O. Carpenter, "Functional significance of cell size in spinal motoneurons," *J Neurophysiol*, 1965, v. 28, p. 560-580.

[2] J. Ame, C. Rivault, J. Deneubourg, "Cockroach aggregation based on strain odour recognition," *Animal Behavior*, 2004, v.68, pp. 793-807

[3] A. Julius, A. Halasz, V. Kumar, G. Pappas, "Controlling biological systems: the lactose regulation system of Escherichia coli," *in 2007 American Control Conference*, 9-13 July 2007 p. 1305-1310.

[4] K. Lerman, A. Martinoli, A. Galstyan, "A Review of Probabilistic Macroscopic Models for Swarm Robotic Systems," in *Swarm Robotics Workshop: State-of-the-art Survey*, pp. 143152, Springer-Verlag, Berlin, 2005.

[5] C. Anderson, "Linking Micro- to Macro-level Behavior in the Aggressor-Defender-Stalker Game," *Adaptive Behavior*, 2004, Vol. 12, No. 3-4, p. 175-185.

[6] A. Martinoli, K. Easton, W. Agassounon, "Modeling Swarm Robotic Systems: a Case Study in Collaborative Distributed Manipulation," *International J. of Robotics Research*, 2004, v. 23, p. 415-436.

[7] B. Selden, K. Cho, H. Asada, "Segmented binary control of shape memory alloy actuators using the peltier effect," *Proc. IEEE ICRA*, 2004, pp. 4931-4936

[8] M. Hafez, M.D. Lichter, S. Dubowsky, "Optimized binary modular reconfigurable robotic devices", *IEEE/ASME Trans. on Mechatronics*, 2003, v. 8, n.1, pp. 18-25

[9] F. Lorussi, S. Galatolo, C. Caudai, A. Tognetti, D. De Rossi, "Compliance control and Feldman's muscle model," *IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics*, 2006, pp. 1194-1199

[10] O. Soysal, E. Sahin, "Probabilistic Aggregation Strategies in Swarm Robotic Systems," *Proc. IEEE Swarm Intelligence Symposium*, 8-10 June 2005, pp.325-332

[11] S. Berman, A. Halasz, V. Kumar, S. Pratt, "Bio-Inspired Group Behaviors for the Deployment of a Swarm of Robots to Multiple Destinations," *in 2007 International Conference on Robotics and Automation*, 1-14 April 2007, p. 2318-2323.

[12] L. Odhner, J. Ueda, H. Asada, "Stochastic Optimal Control Laws for Cellular Artificial Muscles," *in 2007 International Conference on Robotics and Automation*, 1-14 April 2007, p. 1554-1559.

[13] J. Ueda, L. Odhner, S. Kim, H. Asada, "Distributed Stochastic Control of MEMS-PZT Cellular Actuators with Broadcast Feedback," *in 2006 IEEE-RAS International Conference on Biomedical Robotics and Biomechatronics*, 20-22 February 2006, p. 272-277.

[14] J. Ueda, L. Odhner, H. Asada, "Broadcast Feedback of Stochastic Cellular Actuators Inspired by Biological Muscle Control", To appear in *International J. of Robotics Research.*