A New Reinforcement Learning based Automatic Generation Controller for Hydro-Thermal Power Systems

T. P. Imthias Ahamed, P.S. Sastry and P. S. Nagendra Rao
Dept. Electrical Engineering,
Indian Institute of Science,
Bangalore 560012 India
e-mail: *imthi4@rediffmail.com; sastry,nagendra@ee.iisc.ernet.in*

Abstract—Recently, we have proposed [1] a Reinforcement Learning (RL) based approach for designing Automatic Generation Controller for a two area power system, where we have demonstrated the efficacy of the approach on an identical, simple, two area model. The aim of this paper is to demonstrate an alternative RL-AGC design which is simpler. Its effectiveness is demonstrated by considering a hydro-thermal system whose dynamics is more complicated than the system considered in [1].

1. INTRODUCTION

An interconnected power system consists of independent power utilities interconnected using transmission lines. These independent power utilities serving the neads of a geographical area is called an *"area"*. When a bulk load is added or if a generating units trips in any of the area, the frequency of the entire system comes down. Because of the governor action, the generation in each area will increase. However, the governor action alone cannot bring back the frequency to the scheduled value. Moreover, a load change in one of the area, say area A, results in generation change in all the areas, which results in change in tie-line power flow to control area A. The function of the AGC is to bring back the frequency and tie-line flows close to the scheduled values.

There has been considerable research work attempting to propose better AGC systems based on modern control theory [2], [3], Neural Network [4], [5], [6], and Fuzzy systems theory[7], [8]. Most of the approaches treat AGC and the rest of the system to be a single (feedback) controlled dynamical system operating on one time scale, and often use a linearised system model for the design of AGC. In practice, the AGC function is essentially supervisory in nature. Thus the AGC, acts at a slower time scale compared to rest of the power system.

We have shown recently [1] that the AGC problem can be viewed as a stochastic multi-stage decision making problem or a Markov chain control problem. That is, we view AGC as a supervisory controller that decides on the set point for a separate closed loop control system which tries to achieve the set generation. In [1] we have demonstrated the use of Reinforcement Learning(RL) approach for designing an RL controller for the AGC problem on a simple two area system.

0-7803-7651-X/03/$17.00 ©2003 IEEE

The RL based AGC (RL-AGC) presented in [1] has two inputs, Area Control Error (ACE) and rate of change of ACE. In this paper, we propose a new AGC which uses ACE as its only input. We denote such an AGC (with only ACE as the input) by RLC1. In [1], we have considered an identical two area model where each area is represented by a thermal system. However, the two areas need not be identical. Nanda et al. [9] have done investigation using an interconnected hydro-thermal system. In this paper, we investigate the performance of RLC1 for a two area hydro thermal system.

2. RL BASED AGC DESIGN

In the RL approach, the AGC function is viewed as follows. At each instant, k, k=1,2,..., the AGC observes the current '*state*' of the system, x_k, and takes an '*action*', a_k. Let $\mathcal{X}$ denote the set of all possible states, and let $\mathcal{A}$ denote the set of all possible actions. Within this formulation, any AGC algorithm is a mapping from the set of states, $\mathcal{X}$, to the set of actions, $\mathcal{A}$. We call any such mapping, a policy. If the AGC is following a policy π, then $\pi(x)$ denotes action taken by AGC on observing state x.

Learning a 'good' policy (that is, a good AGC) involves obtaining a mapping from $\mathcal{X}$ to $\mathcal{A}$ so that the objective of keeping ACE within a small band around zero is achieved. We call such a mapping an optimal policy, π^*. For the learning system to find the optimal policy, it needs to get a feedback on how the current control law performs. For this we stipulate that whenever the controlled system makes a transition from $x(k)$ to $x(k+1)$ under an action $a(k)$ we get some evaluative feedback signal, called reinforcement, $r(k) = g(x(k), x(k+1), a(k))$. The function $g(x, y, a)$ is part of design of the learning system and it is expected to capture the control objective.

From the above discussion it is clear that to design an RL-AGC one has to decide on the following. (i) What quantities should constitute as 'state', or equivalently what should be the set $\mathcal{X}$? (ii) What is the set of actions $\mathcal{A}$? (iii) What should constitute the g function? In addition to these design choices, we need an algorithm for learning the optimal policy.

For the RL-AGC presented in this paper, the average value of the ACE (ACE_avg) is the only state variable. Since, we are considering RL algorithm which assumes finite number of states, ACE_avg is discretised to finite levels as follows. Let the maximum value of $|ACE|$ for which AGC is expected to act properly be L_A. (We note here that the choice of L_A is not critical for the algorithm.) If it is required to maintain ACE within ϵ_{ACE}, ACE is discretized as follows. ACE values whose magnitudes are less than ϵ_{ACE} are considered as the zero level and the range of ACE values greater than ϵ_{ACE} but less than L_A are quantized to a finite number of levels, M_A (Where $M_A = \lfloor L_A/(2\epsilon_{ACE})\rfloor$) at equal intervals. The quantized value is the midpoint of the interval. All values of ACE greater than L_A are discretized to the maximum level. Similarly, negative values of ACE, less than $-\epsilon_{ACE}$ are discretized to M_A levels. Thus, ACE is discretized to $2M_A + 1$ levels, that is, $x \in \{-M_A(2\delta), \ldots, -2\delta, 0, 2\delta, \ldots, M_A(2\delta)\}$, where $\delta = \epsilon_{ACE}$.

The control action of the AGC is to change the generation set point, ΔP. Since the range of generation change that can be effected in an AGC cycle is known, we can discretise this range to finite levels. Choosing the permissible range in ΔP as $-U_{MAX}$ to U_{MAX} and minimum step size in ΔP as ΔP_{min}, the action set $\mathcal{A}$ will contain the following $2M_P + 1$ levels, $\{-U_{MAX}, \ldots, -\Delta P_{min}, 0, \Delta P_{min}, \ldots, U_{MAX}\}$. (where $M_P = U_{MAX}/\Delta P_{min}$)

The qualitative objective of the controller is captured in the immediate reinforcement function (g function). As the control objective here is to keep the magnitude of ACE less than ϵ_{ACE}, so, whenever the next state is 'desirable' (i.e. $|ACE_{k+1}| < \epsilon_{ACE}$ which implies $x_{k+1} = 0$) then $g(x_k, x_{k+1}, a_k)$ is assigned a value zero. When the next state is 'undesirable', (i.e. $|ACE_{k+1}| > \epsilon_{ACE}$) then $g(x_k, x_{k+1}, a_k)$ is assigned a value -1. Thus function g (immediate reinforcement function) can be defined as follows;

$$
\begin{aligned}
g(x_k, x_{k+1}, a_k) &= 0 \quad \text{if } x_{k+1} = 0 \\
&= -1 \quad \text{otherwise}
\end{aligned} \tag{1}
$$

For learning the optimal policy, we use an iterative algorithm which is based on Q-learning [10]. This involves learning the so called optimal Q-values or state-action values for all state-action pairs (which is denoted as $Q^*(x,a)$). The optimal policy π^* can be obtained from Q^* using the relation $\pi^*(x) = \arg\max_{a\in\mathcal{A}} Q^*(x,a)$. (See [1] for more details). Thus, to learn an optimal policy π^*, it is enough to learn $Q^*(x,a)$ for all (x,a) pairs.

To find $Q^*(x,a)$ we use the following iterative algorithm. Suppose we have a sequence of samples $(x(k), a(k), x(k+1), r(k))$, $k = 1, 2, \ldots$ and let Q^k be the current estimate of Q^*. Then, Q^{k+1}, the next estimate of Q^* is obtained as

$K_p = 120Hz/pu$	$T_1 = 48.7$s
$T_P = 20$s	$T_2 = 0.513$s
$T_g = 0.08$s	T_R= 5s
$T_t = 0.3$s	T_w= 1s
$R = 2.4Hz/pu$	$T_{12} = 0.545$

Table 1. System parameters for the hydro thermal model

$$
\begin{aligned}
Q^{k+1}(x(k), a(k)) &= Q^k(x(k), a(k)) \\
&\quad + \alpha[g(x(k), x(k+1), a(k)) \\
&\quad + \gamma \max_{a'\in\mathcal{A}} Q^k(x(k+1), a') \\
&\quad - Q^k(x(k), a(k))] \\
Q^{k+1}(\overline{x}, \overline{a}) &= Q^k(\overline{x}, \overline{a}) \; \forall(\overline{x}, \overline{a}) \neq (x(k), a(k)),
\end{aligned} \tag{2}
$$

where $0 < \alpha < 1$ is a constant called the step size of learning. If α is sufficiently small then the above iterative algorithm will result in Q^k converging to Q^* if all possible (x,a) combinations of state and action occur sufficiently often in our sample sequence[11], [10].

3. Two Area System

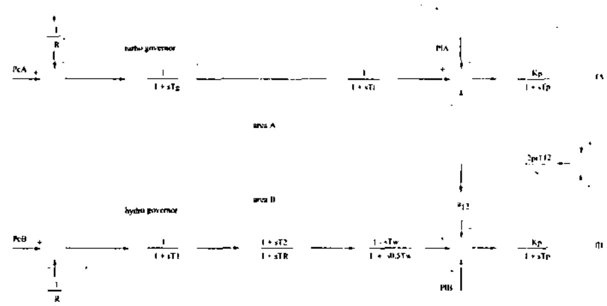

Figure 1. The Two Area Model of the Hydro Thermal System

In this section, we study the performance of RLC1 using a two area hydro-thermal system. In a hydro-thermal system the two areas have widely different characteristics. For the simulation studies presented in this section, we use the hydro-thermal model given in [9]. A block schematic of the model is given in figure 1 and the parameters of the model are given in Table 1.

As mentioned in section 2, to implement a particular RLC1,

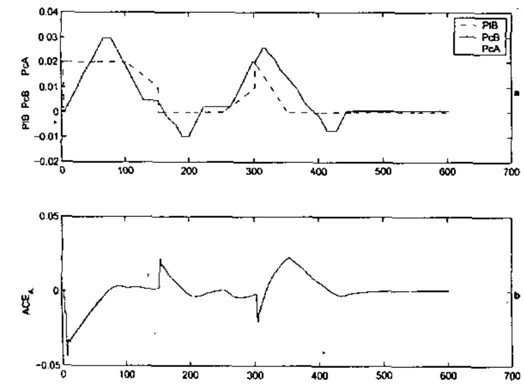

Figure 2. Response of the hydro area with the RL Controller
(a) Plots of load disturbance in area B(PlB), change in set point to governor in area B (PcB) and change in set point to governor in area A (PcA)
(b) Plot of average value of Area Control Error in area B (ACE_B)
X-axis shows time in seconds and all quantities in Y-axis are in p.u.

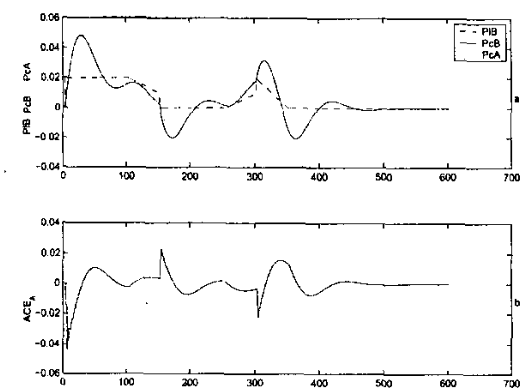

Figure 3. Response of area B with the integral controller

we have to choose $L_A, \epsilon_{ACE}, M, U_{MAX}$ and ΔP_{min}. We have chosen $L_A = 0.02, \epsilon_{ACE} = 0.002, M = 11, U_{MAX} = 0.001$ and $\Delta P_{min} = 0.0005$. The guidelines for choosing these parameters can be found in [12].

Using these parameters, the optimal policy is learned using the simulation scheme described in [1]. We give the response of the hydro area (area B), with sequence of load disturbances in area B and without any load disturbance in area A, in figure 2. To compare the performance, the response of the hydro-thermal system for the same disturbance but with an integral controller whose gain $K_I = 0.15$ has also been obtained. The response of area B with the integral controller for the same load variation is given in figure 3.

From sub-plot a of figure 2 we see that PcB follows the load change in area B. However, there is a large overshoot for the step change in load in area B at t = 5s. PcB increases up to 0.03. A similar large overshoot can be seen in the output signal of the integral controller in sub-plot a of figure 3; PcB increases up to 0.045. By comparing the overshoots with the RL controller and the integral controller, we see that the overshoot with the RL controller is much less. Both the RL controller and the integral controller follow the gradually decreasing ramp (t= 100s to t =150s), the gradually increasing ramp (t= 250s to t =300s), and the decreasing ramp (t= 300s to t =350s). Though the decreasing ramp reduces the load change (PlB) to zero at t=350s, PcB (with RLC1) settles down to zero only at t = 450s (refer figure 2). But PcB for the integral controller oscillates and settles down to zero only around t = 500s (refer figure 3). From the sub-plot b of figure 2 and figure 3, we see that both RLC1 and the integral controller try to bring back the ACE to zero. From the plot of ACE from 400s to 500s in figure 2 and figure 3, it appears that settling time of the ACE with RLC1 is less than the settling time of the ACE with the integral controller. We have seen that the RLC1 response of the area A (thermal) is very good. It must be pointed out here that the observed performance of the RLC1 for areas having hydro-units is quite natural. The inherent dynamics of the hydro unit model (just the primary response to step changes) is very sluggish (settling times of the order of 100s) and oscillatory. However, the superiority of the proposed controller over the integral controller is evident.

4. Conclusions

In this paper, we have presented a new RL-AGC, RLC1, which uses ACE as the only state variable. Further, we have studied the performance of RLC1 for a hydrothermal system where two areas have widely different characteristics. RLC1 in both areas could bring back ACE within a specified bounds. The results here clearly establish the feasibility of using only one state variable for RL-AGC design. This alternative AGC design reduces both the design and implementation complexities.

References

[1] T. P. I. Ahamed, P. S. N. Rao, and P. S. Sastry, "A reinforcement learning approach to automatic generation control," *Electric Power Systems Research*, vol. 63, pp. 9–26, 2002.

[2] M. L. Kothari, J. Nanda, D. P. Kothari, and D. Das, "Discrete mode AGC of a two area reheat thermal system with new ACE," *IEEE Transactions on Power Systems*, vol. 4, pp. 730–738, May 1989.

[3] R. R. Shoults and J. A. Jativa, "Multi area adaptive LFC developed for a comprehensive AGC simulator," *IEEE Transactions on Power Systems*, vol. 8, pp. 541–547, 1993.

[4] Q. H. Wu, B. W. Hogg, and G. W. Irwin, "A neural network regulator for turbo generator," *IEEE Transactions*

on *Neural Networks*, vol. 3, pp. 95–100, 1992.

[5] D. K. Chaturvedi, P. S. Satsangi, and P. K. Kalra, "Load frequency control: a genaralised neural network approach," *International Journal of Electric Power and Energy Systems*, vol. 21, pp. 405–415, 1999.

[6] H. L. Zeynelgil, A. Demiroren, and N. S. Sengor, "The application of ann technique to automatic generation control for multi-area power system," *International Journal of Electric Power and Energy Systems*, vol. 24, pp. 345–354, 2002.

[7] J. Talaq and F. Al-Basri, "Adaptive fuzzy gain scheduling for load frequency control," *IEEE Transactions on Power Systems*, vol. 14, pp. 145–150, 1999.

[8] C. Indulkar and B. Raj, "Application of fuzzy controller to automatic generation control," *Electric Machines and Power Systems*, vol. 23, pp. 209–220, Mar. 1995.

[9] J. Nanda, M. L. Kothari, and P. S. Satsangi, "Automatic generation control of interconnected hydrothermal systems in continuous and discrete mode considering GRC," *IEE Proceedings - part D*, vol. 130, Jan. 1983.

[10] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA, 1998.

[11] D. P. Bertsekas and J. N. Tsitsikilis, *Neuro Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[12] T. P. I. Ahamed, *Reinforcement Learning Based Approaches to Automatic Generation Control.* PhD thesis, Indian Institute of Science, Bangalore 560012, India, 2001.