Volume 3, Issue 10, October 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Optimal Goals Allocation in the Collective of Autonomous Robots Control Problem

Akzhalova Assel*, Satiev Amirzhan, Zhumabayev Akylbek

Computer Engineering Department

Kazakh-British Technical University

Kazakhstan

Abstract— The one of most important challenges within fully autonomous robots research is to introduce a collective of autonomous mini robots control. We consider the autonomous control of collective robotic system that addresses an optimal allocation of goals between robots problem in non-deterministic environment to reach certain global goal. This paper offers the framework of autonomous control system for the group of robots and a selection algorithm that intends to solve above problem employing "swarm" intelligence and reinforcement learning approaches basing on degree of quality performance of the robot updated by equipped real-time analyzer. This hybridization provides a certain reliability level as it makes a whole robotic system be sustainable depending on dynamically changing external environment conditions and flexibility as it takes into account quality degree of each robot.

Keywords— Collective robots, autonomous, optimal control, swarm, reinforcement learning

I. INTRODUCTION

Collective control of autonomous mini robots usually is build over the infrastructure of the system of robots, series of intelligent devices such as sensor-actuators that acquire necessary information and detect objects within a wide range and able to interact with other devices through the communication network. This kind of collective robotic system autonomously has to make decision in order to complete one global task or find best solution that allows reaching common target by intelligent distributing sub-targets while the requirements for time completion and performance should be satisfied. Autonomous control of robotic system has certain advantages as it is efficient and does not require manual control all the time. However, one of the most challenging problems in developing autonomous control of collective robotic system must be addressed to designing intelligent approach of the optimal allocation of goals between robots in non-deterministic environment to reach certain global goal. An allocation of goals between robots in the real life traditionally can be met in any robotic battle systems. The most powerful system is that causing maximum damage to all specified enemy targets. It can be achieved by assigning sub-groups for some goals and keeping strategy for the whole group of robots at each time step. Therefore, group and goal matching selection can be considered as a control function to maximize total value of the damage which is cost function or value of performance of the system. When the magnitude of the damage reaches a maximum corresponding distribution can be logically considered as optimal.

The problem of optimal goal and robot matchmaking can be solved by using special heuristic search algorithms that can significantly speed up the selection process by eliminating unlikely choices. The results of implementing heuristic search algorithms usually differ slightly from the optimal solution although the real-time systems using this approach may consider it as one of the most applicable as the existing mini robots have limited resources, less powerful processors and communication frequency affects dramatically on the performance. Having these constraints and taking into account that the optimal decision should be done in the uncertain environment, the goal and robot selection algorithms should be fairly simple but capable to provide striking strategy solving the global problem.

One of the most interesting approaches that solve specified problem is based on smart technology of collective splitting tasks [1]. Intelligent distribution of tasks techniques between the robots are presented in a number of publications that are based on the ant colony optimization algorithms [2, 3, 4, 5], or the principles of reinforcements learning [6, 7, 8]. Authors in [9] propose a self-learning control track method, where the core of the approach is an approximate dynamic programming. This method improves performance of the supervisory machinery robotic system having only approximate information about its vehicle dynamics. The proposed method is based on reinforcement learning algorithm, the so-called iterative algorithm with least squares kernel (Kernel Least-Squares Policy Iteration-KLSPI). [10] offers a multi-tiered infrastructure of double "collective mind" with three channels of communication that inherit traditional technologies "collective mind" for the effective interaction between multiple layers. This approach improves the manageability of the system by introducing a new entity type called " virtual entity " and new control strategies.

[11] propose an adaptive control based on reinforcement learning for two-level adaptive manipulator. Implemented in [11] decentralized control mechanisms for the multi-component systems are based on the method of linear-quadratic regulator (LQR) and adaptive dynamic programming. This approach ensures effective control by varying the weights of the different loading of the system in real time.

This paper develops a framework of the autonomous collective robotic control system and a selection mechanism that solves the problem of completing global goal (defeating threats) by choosing best candidates for each specific goal. This is achieved by employing "swarm" intelligence and reinforcement learning approaches. This hybridization provides a certain reliability level as it makes a whole robotic system be sustainable depending on dynamically changing environment.

II. A FRAMEWORK OF AUTONOMOUS CONTROL COLLECTIVE OF ROBOTS

We consider an autonomous mini robots control system that consists of the following components: a network of robots that can perceive and response to the environmental changes using sensors and effectors providing collective performance of the global task via achieving specified goals. Figure 1 shows an infrastructure of the control system.

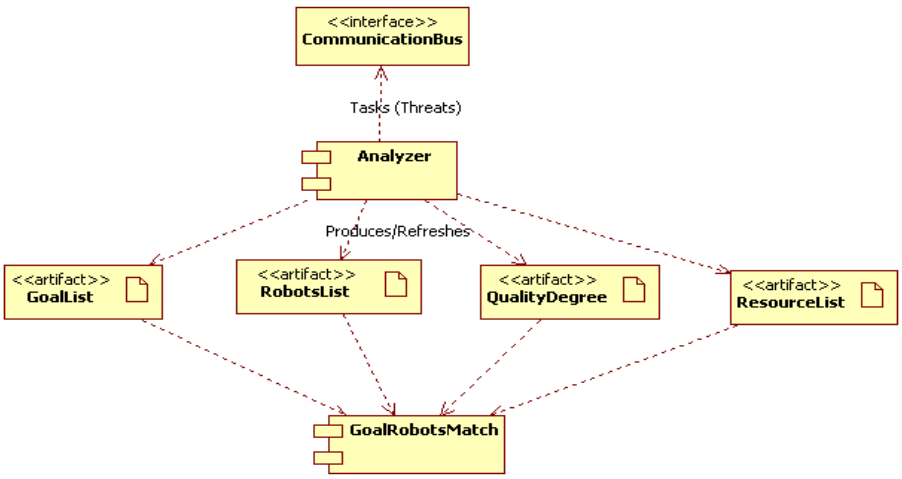

Fig. 1 Goal and Robot Matching control framework

All data are read from *CommunicationBus* and it also used to transmit or receive messages from neighbours of the robot in the group. After analysing data *Analyzer* inserts new goal or removes implemented one from *GoalList* and similarly constraints to/from *RobotList*. These are main artifacts that are used in the formulation of optimization problem to find best match between appropriate robots and goals. *GoalList* essentially represents a list of goals that should be achieved. We do not specify how data transferred from one task to another one, however, it can be represented as message flows or queues. *ResourceList* is formed basing on the information about processor CPU, memory load, power supply that all together represent requirements for the system. On other hand, all robots diverse from each other the quality of the goal performance which can be qualitatively and quantitavely evaluated. as soon as task will be completed. New tasks (threats) will be monitored by *Analyzer* and it is responsible for updating all lists. Therefore, *Analyzer* plays important role in making best choice to perform the global task if new goal matches with previous one. *QualityDegree* is updated by *Analyzer* and it creates or refreshes the degree of Robot for specific task (threat). *QualtyDegree* affects both to the objective function and constraints list.

In next section we offer an approach to match robots and goals in order to execute the global task according to requirements formulated by *Analyzer*.

III. A FRAMEWORK OF AUTONOMOUS CONTROL COLLECTIVE OF ROBOTS

Assume that there is a group of n communicating robots. Suppose there are m goals that must be distributed among the robots in the group. We introduce a vector-function $Q(t)$ for each robot of the group where $q_{ij}^{k}(t)$ determines the amount of damage inflicted on the enemy by robot *i* when it achieves the *j*-th goal to cope the threat *k*. In particular, q_{ij}^{k} can be represented as a degree of threat k danger for the robotic system. This degree of danger can be determined by expert way in the assumption, that all threats form the full group of events, i.e.

$$0 \le q_{ij}^{k}(t) \le 1;\ \sum_{k=1}^{l} q_{ij}^{k}(t) = 1$$

where *l* is an overall number of observed threats.

Every threat is characterized by probability of its occurrence P_i^k . The occurrence probability of threat P_i^k is determined statistically and corresponds to relative frequency of its occurrence:

$$P_i^k = \frac{\lambda}{\sum_{k=1}^{l} \lambda_k} = \overline{\lambda} \qquad (1)$$

where λ_k - frequency of occurrence of k.

The autonomous control system is able to defeat threats fully or partially. The quality of the robot i performance is evaluated by the probability of elimination of every threat k as $P_{ij}^{k\_removal}$.

Let the probability $P_{ij}^{k\_removal}$ is represented by some function evaluating quality of defeating threat:

$$P_{ij}^{k\_removal} = f^k(x_{ij}^k) \tag{2}$$

where x_{ij}^k is the degree of performance by the robot i to eliminate threat k by achieving j-th goal.

The quantitative value of the degree of the performance by robot i-th will be defined by its closeness to desirable result. In order to assess x_{ij}^k we use its normalized value as follows:

$$\overline{x_{ij}}, (j = \overline{1,k}), 0 \le x_{ij} \le 1$$

$$\bar{x}_{ij}^k = \frac{x_{ij}^k - xbest_{ij}^k}{xbest_{ij}^k - xworst_{ij}^k} \tag{3}$$

where $xbest_{ij}^k, xworst_{ij}^k$ are the best and worst values.

Function $f^k(x_{ij}^k)$ can be considered as a membership function that maps each element x_{ij}^k to some number from an interval (0,1) describing a degree of the belonging of element x_{ij}^k to fuzzy set (Table 1).

TABLE I
SCALE FOR QUALITATIVE AND QUANTITATIVE DEGREE

	Threat danger:					
Quality standard (value of a linguistic variable)	rather seldom	more - less seldom	neither frequently	nor it is rare	more – less frequently	Rather frequently
Quantitative value	0.1	0.3	0.5	0.7	0.9	1
	Damage:					
Quality standard (value of a linguistic variable)	very small	Small	average	big	very big	
Quantitative value	0.1	0.3	0.5	0.7	0.9	
Intermediate values between the next ratings	0.2, 0.4, 0.6, 0.8					

We introduce an intelligent control system that reduces an overall damage *W* by achieving all goals to defeat all threats at each time step *t* . As the collective robots have to conduct their work in non-deterministic environment we within an interval $[t, t+\Delta t]$ when the system will reconfigure its state over time Δt from one state to another. We designate a general prevented damage by $\overline{W}(t)$ and the prevented damage due to elimination the threat as a vector $\overline{w}(t)$.

We can express an overall prevented damage as some function depending on probability of the threat occurrence and its danger as follows:

$$\overline{W}(t) = F(t, P_{ij}^k, q_{ij}^k(t), P_{ij}^{k\_removal}; k = \overline{1,l}) \tag{4}$$

and the prevented damage due to defeating threat *k*

$$\overline{\omega}(t) = P_{ij}^k \cdot q_{ij}^k(t) \cdot P_{ij}^{k\_removal} \cdot \Delta t, \quad k = \overline{1,l}. \tag{5}$$

In other words, taking into account that threats independent and additive form of their consequences we rewrite an overall prevented damage as:

$$\overline{W}(t) = \sum_{k=1}^{l} P_{ij}^k \cdot q_{ij}^k(t) \cdot P_{ij}^{k\_removal} \tag{6}$$

Applying (1) and (3) we transform (6) as follows:

$$\overline{W}(t) = \sum_{k=1}^{l} \sum_{j=1}^{m} \overline{\lambda_k} \cdot q_{ij}^k(t) \cdot f^k(x_{ij}^k) \cdot \bar{x}_{ij}^k \tag{7}$$

Now we can formulate a problem of allocation goals in the group of robots as an optimal control problem. It is necessary to choose the best matching goals and robots providing a maximum of prevented damage from threats with desirable expenses.

The formal statement of the problem has the following form:

We need to find

$$\max \overline{W}(t) = W^* \quad x_{ij}^k; i = \overline{1,n};\ j = \overline{1,m}, k = \overline{1,l} \tag{8}$$

which is subject to constraints:

$$C(x_{ij}^k) \le C_{des},\ \ i = \overline{1,n},\ \ j = \overline{1,m},\ \ k = \overline{1,l} \tag{9}$$

where C_{des} desirable cost of the system and $C(x_{ij}^k)$ given cost of chosen robot *i* to implement goal *j* to defeat threat *k*. Table 2 and 3 show example of assigning cost and degree of quality performance by robot *i*.

Table II
Cost of expenses on the damage prevention by robot *i*

R* / T*	1	2	3	4	5	6	7	8
1	0.6	0.7	0.6	0.3	0.6	0.6	0.2	1
2	0.7	0.6	0.6	0.7	0.5	0.4	0.8	0.9
3	0.4	0.3	0.4	0.2	0.3	0.2	0.1	0
4	0	0	0	0.4	0	0.3	0.2	0.1
5	0.2	0.2	0	0.2	0.25	0.2	0	0
6	0.6	0.5	0.6	0.4	0.3	0.3	0.3	0
7	0.5	0.5	0.5	0.7	0.5	0.5	0.7	0
8	0.8	0.7	0.9	0.9	0.8	0.8	0.5	0.9

R* - number of goals; T* - number of threats

Table III
The degree of quality performance by robot *i*

$f(x_{ij})$ / T	0.9	0.8	1.0	1.0	0.8	1.0	0.8	0.9
1								
2		0.1						
3			0.3			0.3	0.3	0.3
4								
5	0.5			0.5	0.5			
6				0.3		0.3	0.1	0.2
7	0.2	0.5	0.3	0.0	0.2		0.3	0.2
8	0.1		0.1	0.0	0.1	0.1	0.1	0.1

T - number of threats

We can consider the formulated problem (8), (9) as a problem of global unconstrained minimization of the objective function $\overline{W}$ if we put our constraints inside to the evaluation function *I*:

$$\min_{X \in R} I(X) = I^*(X) \tag{10}$$

where

$$I = W(X), X = (t, \bar{x}, C, Q, \overline{P}, \overline{\lambda})$$

where *X* is a control function.

According to PSO [12], the robots can be represented as particles in the parameter space of the optimization problem. At each time step the particles G are characterized by their position in a search space *Pos* and velocity vector V. For each position of the particle we calculate value of the evaluation function and depending on the found value and certain rules particle changes its position and velocity of the search space.

We use PSO to calculate the best value for the evaluation function *I*. However, in order to memorize best solution and provide intelligent way of collective control of robot-agent we employ Q-learning approach [17]. The Q-Learning algorithm allows choosing between immediate rewards and delayed rewards as well as include punishment mechanism.

The transition rule for the system of robots according to the Q-learning algorithm can be formulated as follows:

*Q(state, action) = Rf(state, action) + gamma * Max[Q(next state, all actions)]*

where $Rf(state, action)$ reinforcement of conducting action that transforms the system from one state to another one; *gamma* is a positive number less than 1, and ensures the convergence of the sum *Q*. Learning process for the robot can be considered as finding the sequence of actions which maximizes the sum of the future reinforcements and, therefore, it gives the shortest path to the best solution.

We have developed simulation tool that is based on proposed approach. The results of simulation visualized on Figure 1 and 2 where Figure shows initial state of the system and Figure 2 shows last state of the system after 2000 iterations. For the experiments we use 100 agents and 5 threats which are fire sources (blue line means obstacle).

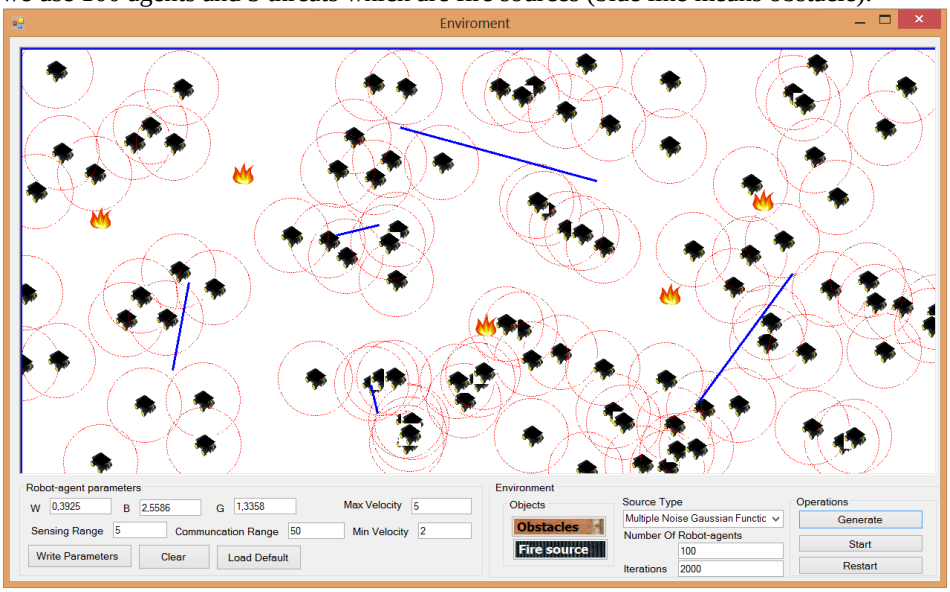

Fig. 2 Initial state of the Goal and Robot Matching control framework

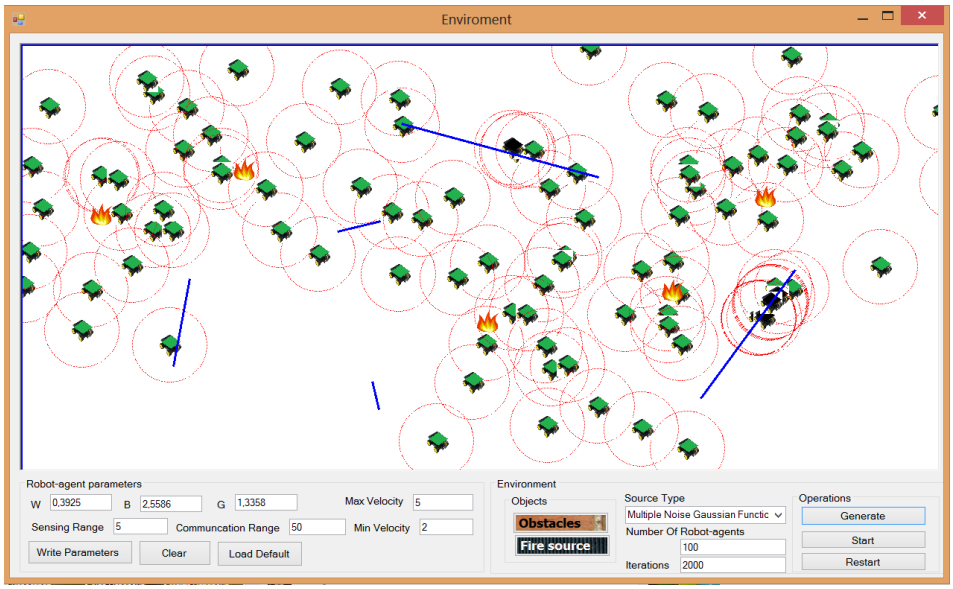

Fig. 3 Result after simulation of the Goal and Robot Matching control framework

IV. CONCLUSIONS

This paper considers the framework maintains three main requirements for the collective robotic autonomous control system: analyze data and threats, uses experience of the problem solving by employing Q-learning approach, and every time finds an optimal robot and goal matching to defeat the threat basing on degree of quality performance of the robot and taking into account non-deterministic environment. In addition, we achieve time reduction by solving optimal control problem by applying PSO technique.

ACKNOWLEDGMENT

This work is carrying out under the grant (*N73 04.02.2013*) of Ministry of Education and Science of Republic of Kazakhstan funding research project at Kazakh-British Technical University ”Development of intelligent autonomous control technology of mobile multi-component robotic systems”.

REFERENCES

[1] A.M. Khamis, M.S. Kamel, M.A. Salichs, Cooperation: concepts and general typology, in: IEEE International Conference on Systems, Man and Cybernetics, vol. 2, 2006, pp. 1499–1505

[2] A. Chatty, I. Kallel, A.M. Alimi, Counter-ant algorithm for evolving multirobot collaboration, in: Proceedings of the 5th international Conference on Soft Computing As Transdisciplinary Science and Technology, 2008, pp. 84–89.

[3] Y. Ding, M. Zhu, Y. He, J. Jiang, An autonomous task allocation method of the multi-robot system, in: International Conference on Control, Automation, Robotics and Vision, 2006, pp. 1–6.

[4] T. Zheng, L. Yang, Optimal ant colony algorithm based multi-robot task allocation and processing sequence scheduling, in: 7th World Congress on Intelligent Control and Automation (WCICA 2008), 2008, pp. 5693–5698.

[5] Y. Zou, D. Luo, A modified ant colony algorithm used for multi-robot odor source localization, LNCS 5227 (2008) 502–509.

[6] Y. Takahashi, K. Edazawa, K. Noma, M. Asada, Simultaneous learning to acquire competitive behaviors in multi-agent system based on a modular learning system, in: IEEE/RSJ Intelligent Robots and Systems (IROS 2005), 2005, pp. 2016–2022.

[7] E. Yang, D. Gu, Multiagent reinforcement learning for multi-robot systems: a survey, Tech. Report, Univ. Essex, 2004

[8] C. Zhou, Robot learning with GA-based fuzzy reinforcement learning agents, Information Sciences 145 (1-2) (2002) 45–68.

[9] Xin Xu , Hongyu Zhang ; Bin Dai ; Han-gen He Self-learning path-tracking control of autonomous vehicles using kernel-based approximate dynamic programming, Page(s): 2182 - 2189 Proceedings of IEEE World Congress on Computational Intelligence. Neural Networks, 2008. IJCNN 2008

[10] Li, Wenfeng and Shen, Weiming Swarm behavior control of mobile multi-robots with wireless sensor networks // Journal of Network and Computer Applications archive Volume 34 Issue 4, July, 2011, Pages 1398-1407

[11] Bidyadhar Subudhi, IEEE and Santanu Kumar Pradhan Direct Adaptive control of a flexible robot using reinforcement learning // Proceedings of the Industrial Electronics, Control & Robotics (IECR), 2010 International Conference, pp.129-136

[12] R. Mendes, J. Kennedy, J. Neves. The fully informed particle swarm: Simpler, maybe better. // IEEE Transactions on Evolutionary Computation. - 2004, v. 8, pp. 204 - 210.

