Autonomic Learning Through Stochastic Games for Rational Allocation of Scarce Medical Resources

Rossi Kamal and Choong Seon Hong
Department of Computer Engineering
Kyung Hee University
Email:(rossi and cshong)@khu.ac.kr

Mi Jung Choi
Department of Computer Science
Kangwon National University, Korea
Email: mjchoi@kangwon.ac.kr

***Abstract*—The confront of medical ethics, 'rational allocation of scarce medical resources' practically appears as uncertainty in secondary-spectrum-usage in a cognitive Body Area Network(BAN). In this context, we have formulated RMSRA-VB(*R*ational *M*edical *S*carce *R*esource *A*llocator *V*irtual *B*ackbone) as an initial step to provide autonomy in wireless agents so that they can adapt to that uncertainty caused by collision-avoidance and EMI(Electro-Magnetic Interference)-avoidance. Then, we have proposed a distributed autonomic learning framework, RMSRA-QL-POSG(RMSRA using *Q-L*earning in a *P*artial *O*bservable *S*tochastic *G*ame Model). RMSRA-QL-POSG is the first Q-learing algorithm in a POSG Game Model. By RMSRA-QL-POSG, wireless agents learn to utlize secondary-channel in uncertain cognitive BAN. Our probabilistic analysis proves how RMSRA-QL-POSG is successful in inferring uncertainty , in terms of value/reward-over-belief calculation in time horizons, T=1 and T=2, by considering wireless agents' states(monitoring, emergency) and observations (EMI-effect and collision-avoidance).Proof of convergence and complexity analysis of RMSRA-QL-POSG are also presented.**

I. Introduction

Ever since first appeared in medical ethics anthologies, the slogan 'Who Shall Live When Not All Can Live?' has become an important issue in 'rationality of scarce medical resource allocation', which can be based on any of principles like 'equal to everyone', 'favoring the worst-off', 'maximizing total benefits' or 'promoting social usefulness', etc[1]

A. Rational Medical-Scarce-Resource Allocator Virtual Backbone for Uncertainty Formulation

Such is the case in a medical body area network[2](Fig.1). However, autonomic management aims to provide self-learning capacity to M2M-based wireless healthcare nodes, such that it can adapt to uncertainty by learning network-context. In this context, we formulate RMSRA-VB (m connected k-dominating set with multiple uncertainty constraints (EMI, collision, etc.)) (Fig.2) as an initial step for autonomy under uncertain communication in a cognitive BAN (Fig. 1). So, Autonomic Manager-Gateway and Gateway-Mobile Agent connections can tolerate up to $m-1$ and $k-1$ communication uncertainty, respectively Let us have an example how RMSRA-VB reasons uncertainty in a cognitive BAN. Mobile Agent (A2) at patient with normal condition senses his vital sign and at T=1, sends to Gateway G1, other than G2 for the uncertainty of EMI-effect to a combined cancer patient. At T=2, G1 waits because there is a uncertainty of spectrum collision, because Telemedicine unit and A4 at emergency patient are occupying primary and secondary channels. So, at T=3, G1 sends monitoring information of A2 to Autonomic Manager. We consider that RMSRA-VB can be constructed by greedy algorithm like [5].

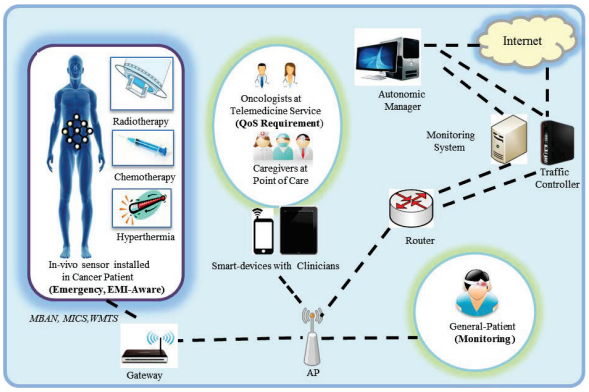

Fig. 1. Conceptual Model-Autonomic Learning in a M2M-based Smart Health Environment for Rational Allocation for Scarce Medical Resources

B. Autonomy under Uncertainty by Unique POSG-QL based Distributed Learning Framework

As wireless network deployed in a hospital environment is a stationary mesh network, wireless nodes observe changes in surrounding environment, because of continual operation, electromagnetic interference, etc. We have formulated as if multiple agents are working in a stochastic game where the play proceeds by steps from position to position, according to transition probabilities controlled jointly by multiple wireless agents. Considering the uncertainty of cognitive BAN, we have formulated it as partial observable stochastic model that suits the best to it. Therefore, we have proposed unique distributed approach RMSRA-QL-POSG(*R*ational *M*edical *S*carce *R*esource *A*llocation using *Q-L*earning in a *P*artial *O*bservable *S*tochastic *G*ame *M*odel). To our knowledge, RMSRA-QL-POSG is the first multi-agent Q-learning algorithm in partial observable stochastic game model, whereas Hansel et al.'s POSG-DP[3] is a dynamic algorithm in the same game model.

C. Effective Uncertainty Inferring by RMSRA-QL-POSG

By considering that wireless agent can be either in regular/monitoring or emergency state in a cognitive BAN, our

978-3-901882-50-0 ©2013 IFIP

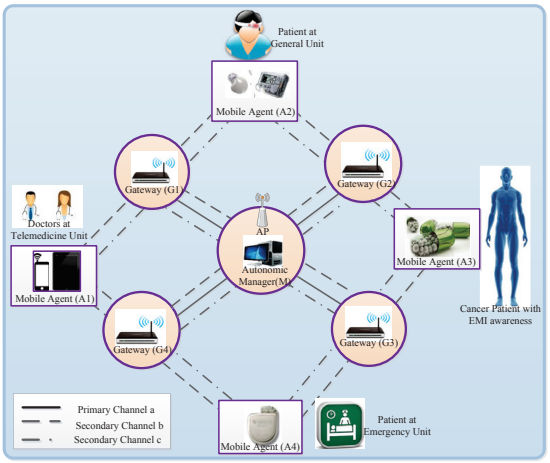

Fig. 2. Uncertainty Formulation with RMSRA-VB

TABLE I. DESCRIPTION OF RMSRA-QL-POSG

Multi-Agent Q-Learning Properties	**How $RMSRA-QL-POSG$ is applied?**
Agent	Cognitive BAN nodes.
Policy (π)	Rational medical scarce-resource allocation in a Cognitive BAN
Reward (r)	Estimated cost achieved by avoiding EMI and collision in spectrum usage in cognitive BAN
Action (a)	Connect/release link to RMSRA-VB
State(s)	Emergency, Monitoring
Value(Q)	Value of joint observation, joint action pair based on immediate reward and future expected rewards.
Observation (O)	Neighbor node is EMI-sensitive/safe or collision-prone/safe in spectrum usage
Convergence	Converge to Nash-Equilibrium, Min-Max Equilibrium and Correlated Equilibrium(Section 4.3)

probabilistic analysis(Section IV) proves how RMSRA-QL-POSG is successful in inferring uncertainty, in terms of value/reward(over belief) calculation in time horizons, $T = 1$ and $T = 2$. However, the convergence of RMSRA-QL-POSG is proved by showing how policy-upgrading stage converges to correlated Equilibrium, Min-Max Equilibrium and Nash-Equilibrium[4], in three different healthcare scenarios. Moreover, the complexity of RMSRA-QL-POSG is analyzed by comparing with POSG-DP[3], DEC-POMDP[3] and relevant[3] techniques.

II. PROPOSED DISTRIBUTED AUTONOMIC LEARNING FRAMEWORK RMSRA-QL-POSG

A. Proposed Unique POSG-QL-based RMSRA-QL-POSG Algorithm

RMSRA-QL-POSG(Algorithm 1) enables wireless agents, equipped with RMSRA-VB, to learn to utilize secondary spectrum, by avoiding collision or EMI-effect. Table I summarizes how RMSRA-QL-POSG applies this multi-agent Q-learning algorithm. Each iteration involves policy upgrading by using POL-UP-COG-VB (Algorithm 2). We have explained all steps of RMSRA-QL-POSG by example formulation(Fig. 3) in the next subsection.

Algorithm 1 RMSRA-QL-POSG

1. Initialization
2. initialize $P_1(o|s,a)$,$P(s^{'}|s,a)$,$R(o,a)$
3. Iteration
4. **while** Healthcare network is in operation **do**
5. initialize belief state
6. Choose action $\bar{A}$ using policy derived from Q
7. Take action $\bar{A}$, observe $\bar{O}$ and $\bar{R}$
8. Policy Upgrading
9. Upgrade policy using POL-UP-COGVB
10. Upgrade Q as
11. $Q(o,a) = \sum_{s \epsilon S} b(s) R(o,a) +$

$$\gamma \sum_{o \epsilon O} P(o|s,a) P(s^{'}|s,a)) max_{a^{'}} Q(o^{'}, a^{'}) \quad (1)$$

12. $s \leftarrow s^{'}$
13. **end while**

Algorithm 2 POL-UP-COG-BAN

1. **while** Secondary usage of spectrum is possible **do**
2. **if** EMI-sensitive healthcare device is adjacent **then**
3. Base-station controls spectrum access of wireless devices
4. **end if**
5. **if** High-bandwidth demanding telemedicine application is available **then**
6. By considering minimum-level usage with respect to telemedicine application, wireless devices share spectrum to maximize it
7. **end if**
8. **if** No Telemedicine Service or EMI-Effect **then**
9. Wireless devices share spectrum by considering best-responses to each other
10. **end if**
11. **end while**

B. POSG-QL Formulation for Rational Allocation of Scarce Medical Resources

At T (Fig. 3), the optimal value function is obtained recursively by value function at $T-1$. However, as state s is not observable, value function, obtained over belief space at $T = 1$, is

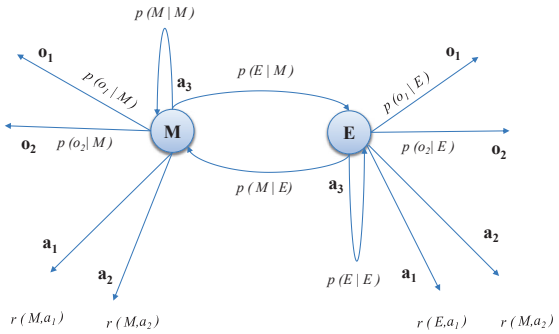

Fig. 3. RMSRA-QL-POSG Formulation

$$V_T(b) = \gamma \max_a [r(b,a) + \int V_{T-1}(b^{'}) p(b^{'}|b,a) db^{'}] \quad (2)$$

$$V_1(b) = \gamma \max_a [E_s(s,a)] \quad (3)$$

$$\pi_T(b) = argmax_a [r(b,a) + \int V_{T-1}(b^{'}) p(b^{'}|b,a) db^{'}] \quad (4)$$

As state is not exactly known, wireless node can either be in monitoring(M) or emergency(E) and can receive reward of either $r(M,a)$ and $r(E,a)$, respectively. So, linear combination of extreme values weighted by probabilities is taken.

$$r(b,a) = E_s[r(s,a)] = p_1 r(M,u) + p_2 r(E,u) \quad (5)$$

So, optimal policy over belief is chosen by considering the action (immediate or delayed packet delivery), which gives the better optimal value over belief.

$$\pi_1(b) = \{^{a_2, if(p_1 \leq x)}_{a_1, if(p_1 \geq x)} \quad (6)$$

$$V_1(b) = max_a r(b,a) = max|r(b,a_1), r(b,a_2), r(b,a_3)| \quad (7)$$

Wireless node can make joint observations (whether neighbor node is EMI-prone or collision-prone), before taking action. However, belief is updated using Bayes rules first. So, given all observations, value over belief can be calculated as follows

$$p_1^{'} = p(M|o_1) = \frac{p(o_1|M)p(M)}{p(o_1)} \quad (8)$$

$$p_2^{'} = p(E|o_1) = \frac{p(o_1|E)p(E)}{p(o_1)} \quad (9)$$

$$V_1(b|o_1) = max|\{r(M,a_1)p(M|o_1)\}, \{r(E,a_1)p(E|o_1)\}| \quad (10)$$

$$\bar{V_1}(b) = E_o[V_1(b|o)] = \sum_{i=1}^{i=2} p(o_i) V_i(b/o_i) \quad (11)$$

Let us consider state-transition for the projection of value function at $T = 2$. Upon executing action $a3$, wireless node can change or keep its state.

$$p_1^{'} = p_1.p(M^{'}|M,a_3) + p_2.p(M^{'}|E,a_3) \quad (12)$$

$$p_2^{'} = p_1.p(E^{'}|E,a_3) + p_2.p(E^{'}|M,a_3) \quad (13)$$

Similarly, value function over belief upon executing $a3$ is calculated by projecting [12][13] to [11].

III. Inferring Uncertainty through RMSRA-POSG-QL

1) Assumptions: Healthcare agent is either at monitoring state(M) or emergency state(E). $P1 = 1$ means that the agent is certainly in monitoring state, whereas $P1 = 0$ means that the agent surely starts with emergency state.

From monitoring state, wireless agent might send information immediately(action $a1$) or wait for some period(action $a2$). Immediate effort might bring spectrum collision($r(M,a1)$ =-100), however, delayed-effort might ensure guaranteed delivery($reward, r(M,a2) = 100$). On the other hand, from emergency state, immediate-effort(a1) is necessary($r(E,a1)$=100), as delayed effort(a2) might result in danger($r(E,a2)$ =-100).

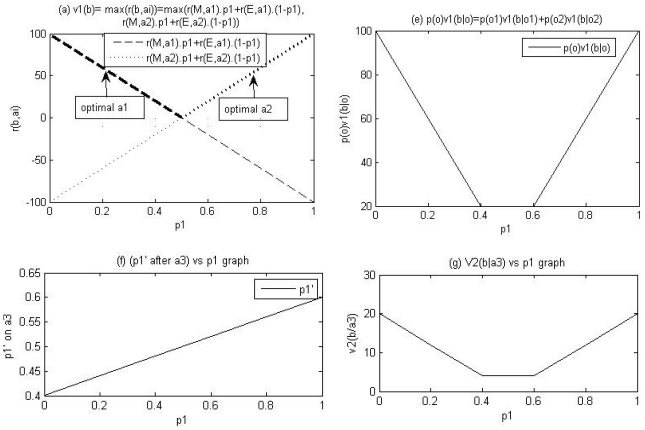

Fig. 4. Immediate Reward, Expected value, State Transition in T=1 and T=2

Healthcare agent uses two observations($o1$,$o2$), before any action, whether it might result in spectrum-collision($o1$) or EMI-effect($o2$), etc. Therefore, we assume $P(o1|M)$=0.6 and $P(o1|M)$=0.4 and $P(o1|E)$=0.4 and $P(o1|E)$=0.6.

Healthcare agent, upon action a3 can change its state(M to E or reverse) or keep its original state. Let, $p1'(M|M) = p1'(E|E) = 0.6$ or $p2'(M|E) = p2'(E|M) = 0.4$.

2) Immediate Reward at Time Horizon,$T = 1$: Fig.4(a) shows that, when healthcare agent is more certainly in emergency state, immediate-effort($a1$) is better action. When agent is more certainly in monitoring state, delayed-effort($a2$) becomes more beneficial.

3) Expected Value at Time Horizon,$T = 1$: Fig.4(e) shows that when agent is less certainly in monitoring state, its expected value over belief(considering joint observations about collision and EMI) reduces gradually at first. In an interim period, this value is unclear (p1=0.4 to p1=0.6). Then, as the agent becomes more certain about monitoring state, this value gradually increases.

4) State Transition and Expected Value at Time Horizon,T=2: Fig 4(f) shows that, the more certainly wireless agent is in monitoring state, the more is the probability to keep its original state (upon action a3) in time horizon T=2. Fig. 4(g) shows the less the agent is certain about emergency state, the less is the expected value. At a certain stage(p1=0.4 to p1=0.6), expected value become stable. However, the more the agent is certain about monitoring state, the more is the expected value.

IV. Performance Analysis of RMSRA-QL-POSG

A. Convergence of RMSRA-QL-POSG

In general condition, all wireless agents compete to use available secondary spectrum by considering best-response to each other. Thus, agents establish Nash-Equilibrium[4] among them. However, when an emergency situation occurs

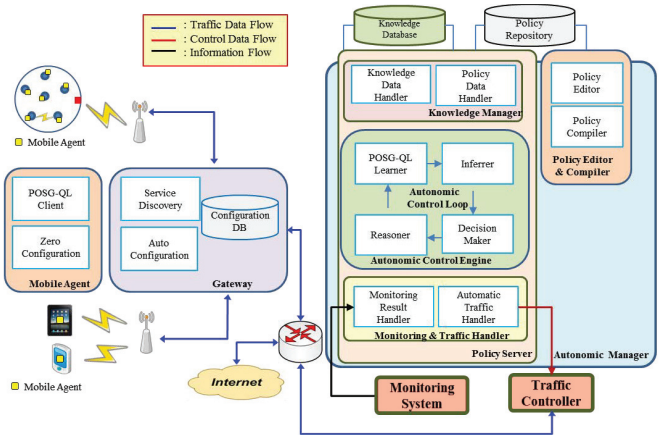

Fig. 5. Autonomic Management Software for M2M Networks in Smart-Health Services

or EMI-effect on a healthcare device becomes acute, base-station(with help of autonomic software) sets up new spectrum allocation strategy for wireless agents. This strategy ensures that wireless agents do not work against emergency or healthcare device, however, agents do not know each other's equilibrium, which is set by base-station. Thus, there exists a correlated equilibrium[4] among agents. Moreover, when there is telemedicine application demanding high-bandwidth, other wireless agents should minimize their spectrum-usage with respect to the telemedicine device. So, wireless agents try to maximize utility in available minimum spectrum. Thus, there exists Min-Max Equilibrium[4] among wireless agents.

B. Comparison of RMSRA-QL-POSG with DEC-POMDP[3],POSG-DP[3],etc.

Approximate solutions for POSG[3] reduce computational complexity by a series of smaller Bayesian games, however the solution is suboptimal. Dynamic programming for DEC-POMDP[3] and POSG[3] considers bottom up dynamic programming, which consume much memory. Improved memory bounded DP for POSG[3], combined top-down heuristic to bottom-up DP to limit the number of policy trees and however it is a sub-optimal solution and there exists a trade-off between number of policy trees and solution quality. However, RMSRA-QL-POSG is the first Q-learning algorithm in POSG model. Likewise any Q-learning algorithm,RMSRA-QL-POSG does not require any environmental model, which is a fundamental requirement in POSG-DP. RMSRA-QL-POSGs advantage is that it can consider both cooperative and noncooperative case, which is not possible in DEC-POMDP[3].

V. Implementation and Future Work

By incorporating POSG-QL-based RMSRA-QL-POSG algorithm, we are developing for last 3 years an Autonomic Management Software[7][6][8][9] for M2M Networks in Smart-Health Services(Fig.5). The uniqueness of the system is Machine Learning-driven Autonomic Control Loop (Reasoner, POSG-QL Learner, Inferrer and Decision-Maker).The system is dedicated to solve two challenges, namely 'autonomous manageability' and 'scalability' of M2M Networks. In this paper, we have considered linear optimization in each phase of Autonomic control loop. In future, we will apply non-linear optimization techniques, especially for reasoning and inference.

Acknowledgment

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2012-0006421). Dr. CS Hong is the corresponding author.

References

[1] G. Persad, A. Wertheimer and E. Emanuel, *Principles for allocation of scarce medical interventions*, The Lancet, vol.373, pp.423-431.2009

[2] J. Wang, M. Ghosh and K. Challapali, *Emerging cognitive radio applications: A survey*, Communications Magazine, IEEE, vol.49, pp.74-81

[3] S. Seuken and S. Zilberstein, *Formal Models and Algorithms for Decentralized Decision Making under Uncertainty*, Autonomous Agents and Multi-Agent Systems, vol.17, pp.190-250, 2008.

[4] L. Busoniu, R. Babuska and B. Schutter, *A Comprehensive Survey of Multiagent Reinforcement Learning*, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol.17, pp.190-250, 2008.

[5] R. Kamal, C. S. Hong, *Fault Tolerant Virtual Backbone for Minimum Temperature in In Vivo Sensor Network*, IEEE Internationnal Conference on Communications(ICC 2012), Ottawa, Canada, 2012.

[6] R. Kamal, M. S. Siddiqui, R.Haw and C.S.Hong, *A policy based management framework for machine to machine networks and services*, 13th Asia-Pacific Network Operations and Management Symposium (APNOMS), Taiwan, 2011.

[7] C.S Hong, R. Kamal, S.H. Shin, S .I. Moon and R. Haw, *Software Design Description-Version 4.0-Policy Based Autonomic Management for M2M Networks and Services*,[Online], Available: http://networking.khu.ac.kr/m2m/sddv4.doc

[8] pbanm-m2m, *Policy-based Autonomic Management Software for Machine to Machine Setworks in E-health Services*,[Online], Available: http://code.google.com/p/pbanm-m2m/

[9] pbnm-m2m, *A Policy-based Network Management Framework for Machine to Machine Networks in E-Health Services*,[Online], Available: http://code.google.com/p/pbnm-m2m/