Decentralized Q-Learning for Uplink Power Control

Sumayyah Dzulkifly [‡], Lorenza Giupponi[*], Fatin Said [‡] and Mischa Dohler [‡]
[‡]Department of Informatics, King's College London, Strand, UK, WC2R 2LS
[*]Centre Technologic Telecomunicacions Catalunya, Av. Carl Friedrich Gauss 7, 08860 Barcelona, Spain
E-mail: [sumayyah.dzulkifly, fatin.said, mischa.dohler]@kcl.ac.uk, lorenza.giupponi@cttc.es

***Abstract*—Fractional power control (FPC) is the simplified version of open loop power control (OLPC) in long term evolution (LTE) that relies on downlink path loss information from base station (BS). This allows user equipment (UE) to decide which power to use for uplink transmission. However, asymmetric behavior of uplink and downlink transmission in crowded network might cause unfair transmit power estimation. This motivates our investigation of implementing uplink path loss and q-learning algorithm to enable UE to decide appropriate transmit power on its own. In this study we apply the concept of FPC into q-learning, enabling UE to find suitable transmit power with respect to uplink path loss. 3GPP uplink path loss model is exploited in our study. We compare outputs between our proposed method and FPC. . From simulation, we find out that DQL performs better as compared to fractional power control in terms of signal-to-interference-noise-ratio (SINR) with average increase factor of 3.5.**

***Index Terms*- fractional power control, open loop power control, decentralized q-learning, long term evolution**

I. INTRODUCTION

Uplink power control is crucial in managing interference and ensuring optimal overall user equipment's (UE) capacity. In long term evolution (LTE), multi-tier network is deployed to supervise dense mobile population. Due to this, managing interference in uplink transmission became more challenging. There are two types of principal uplink power control namely, closed loop power control (CLPC) and open loop power control (OLPC). Full power control is deployed in CLPC where enhanced Node B (eNodeB) stipulate specific transmit power for each UE through control signal. Whereas OLPC has the option of utilizing either full- or fractional- power controls to specify transmit power at UE side [2].

Uplink transmission adopts single-carrier FDMA (SC-FDMA) to ensure battery survival by optimizing the number of symbol per subcarrier. Thus, it is vital to specify the correct transmit power to avoid waste of resources (i.e. battery life). However, there is a need to balance between battery life and ensuring reliable information transmission. Fractional power control is widely utilized in previous studies due to its simplicity and minimal dependency towards eNodeB. UE is able to decide its own transmit power based its own path loss measurement with respect to the serving eNodeB. However, there is instance where UE is near the edge of the cell and in need to transmit at higher power.

Implementing decentralized q-learning (DQL) at UE would enable it to find the most optimal transmit power. DQL is a reinforcement learning that allows UE, identified as agent to learn the wireless environment on its own and share the information with other agent. The learning does not require any prior knowledge which improves the decision making process. Further, there is no main control element in DQL. This enables task distributions among multiple agents. In this study, we integrate the fractional power control from [3] considering path loss and inter-interference from UEs using reward based DQL. This algorithm will be applied in multi-tier LTE. DQL algorithm structure introduced by [4] and [5] will be the reference of our study.

The outline of this paper is as follows; Section II summarizes works related to our study. Section III elaborates our system model. Proposed technique is elaborated in Section IV. Section V discuss and analyze our outputs. Finally, Section VI concludes our study.

II. RELATED WORKS

Investigation on q-learning employment in mobile network to manage resources and reduce interferences is wide and comprehensive. Study by [4] applies cost based multi-agent q-learning to manage interference between macro- cell and femto- cells by manipulating resource block scheduling. In the paper, q-learning algorithm is utilized to enable additional interference information exchange between macro eNode B and femto eNode B through X2 interface. The approach contribute in reducing interference signalling delay between macro- and femto-.

Similarly, [5] utilize reward based q-learning to manage multi-tier interferences transmitting at the same frequency. Meanwhile, [7] apply DQL to manage mutual downlink interference between macro- and femto- cell. The author apply DQL to observe the impact of setting different of transmit power levels on SINR. The most recent works on q-learning carried out by [6] where the authors integrate interactive learning between multi- agents exploiting the same frequency bands in LTE femtocell.

In [10], the author consider traffic management using q-learning where the learning process enables dynamic uplink and downlink association. While authors in [9] employ statistic-based machine learning for uplink power control using cloud mechanism in self-organized network (SON) architecture in LTE. The author model data-driven framework emulating real LTE environment to enable practical implementation in LTE network. Study in [8] propose employment of control parameters to improve open and closed loop power control.

III. SYSTEM MODEL

Our system model is based on simple LTE multi-tier network. In our model, the focus is on uplink transmission where we apply DQL to each UE. We specify maximum UE transmit power, Pmax as 24 dBm as proposed by study in [8]. We assume the path loss measured at UEs as uplink path loss. Our focus is to observe the convergence when DQL is utilized at UE.

For this study, we simplify our model into one macro- and one femto- cell with a uniform number of UEs on each cell. UE in femto- cell is interfered by macro- UE and vice versa during uplink transmission. Fig. 1 illustrate uplink interference scenario. In the figure, UE 2 connected to neighbouring base station (BS) interfere with UE 1 located near the cell edge of serving BS.

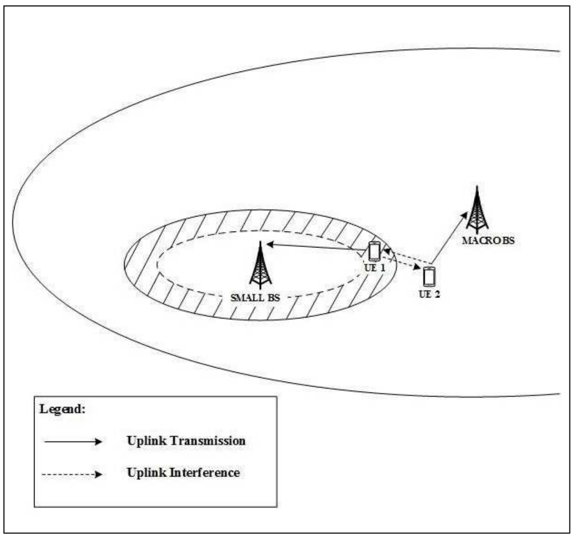

Fig. 1. Uplink Interference Scenario in Multi-tier Networks

A. Open Loop Power Control

Elements from simplified open loop power control, also known as fractional power control (FPC) is utilized in our proposed algorithm as presented by [13] to manage interference during uplink transmission. Generic FPC algorithm, used to specify UE uplink transmission power is as follows:

$$P_{tx,ue} = \min\{P_{max,ue}, P_o + \beta L + \Delta_{ue}\} \quad (1)$$

where $P_{tx,ue}$ is UEs transmit power, P_{max} is the maximum UEs transmit power obtained by distributing maximum power level per number of resource blocks, P_o is cell specific parameter defining the received power per physical resource block (PRB), β is path loss compensation factor which can be set between 0 and 1, L is the estimated UEs path loss and Δ_{ue} is control signal received from eNode B.

If we assume there is no control signal received from eNode B, i.e. $\Delta_{ue} = 0$, the equation is further simplified to:

$$P_{tx,ue} = \min\{P_{max,ue}, P_o + \beta L\} \quad (2)$$

which we are able to manipulate both P_o and β as pointed out in study by [11] [12]. The exploitation of the said variables in equation (2) can be specified based on the system's scenario, namely,

- If UE distribution is more towards the cell edge where UE experience severe interference, higher path loss compensation factor value (i.e. nearing to 1) need to be implemented, considering fixed target signal-to-interference-plus-noise ratio (SINR).
- If UE distribution is mostly near to eNode B, average compensation factor value can be implemented, based on target SINR.

From equation (2), we know that by specifying $P_{max,ue}$ to the highest UE transmit power (i.e. 24 dBm), $P_{tx,ur}$ will specify the minimum transmit power from $P_o + \beta L$. Without exploiting either P_o or β, UE at the cell edge will not be able to transmit effectively due to high interference from other tier UE.

B. Decentralized Q-Learning

Q-learning is a type of reinforcement learning algorithm that allows agent(s) to learn by itself based on added and discounted value when an action is selected. By applying q-learning in our scenario, each agent will be able to exchange information with one another.

From the processing point of view, this would reduce the time consumed and the complexity of the system to sense and analyze its environment before choosing the right action. In other words, each agent(s) involved have the flexibility to learn and react accordingly.

In our model, both power control and SINR is integrated in reward based q-learning where the q-values is updated based on information on state, *c* and action, *d* based on the following equation [5];

$$Q(c,d) \leftarrow Q(c,d) + \alpha * [V^{l} - Q(c,d)] \quad (3)$$

where *c* is a set of q-learning states which determines the q-learning conditions, *d* is a set of possible q-learning actions for decision making, α is the learning factor whom value could range from 0 to 1, and $V^{l} = r + \gamma * \max_d Q(c^{l}, d)$ is the maximum value of the discounted reward received from the next set of state.

IV. DECENTRALIZED Q LEARNING UPLINK POWER CONTROL

This section elaborates the technique used in our study. We deploy multi-tier cells where a number of UEs on each cells is uniformly distributed as shown in fig. 2. The figure illustrate multi-tier region known as macro- cell and small- cell region. In addition to that, there is small- cell edge region where it is crucial for UE to identify appropriate transmit power.

It is assumed that each UEs in the serving cell utilize q-learning algorithm to enable them to select appropriate transmit power with corresponding to their distance to the serving BS. These UE will have updated q-learning table in which information is utilized to enable them to select appropriate transmit power against the interference and the location (i.e. nearing the edge of the cell or near BS).

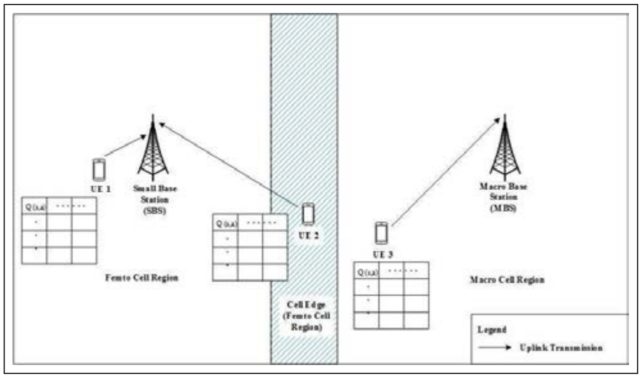

Fig. 2. DQL Uplink Power Control Scenario (Multi-tier Network)

Meanwhile, fig. 3 displays the simulated scenario where some of UEs are close to BS while others are closer to the cell edge. UEs are uniformly distributed within both macro- and small- cell. UE random positions in simulation allow efficient DQL implementation which will be elaborated later. The simulations feature involves in our system model is described in Table 1.

As mentioned before, UEs in both macro- and small- cell adopt q-learning algorithm, enabling them learn uplink path loss information and select the right power level action. In our proposed model, equation (2) is integrated in reward based q-learning equation (3).

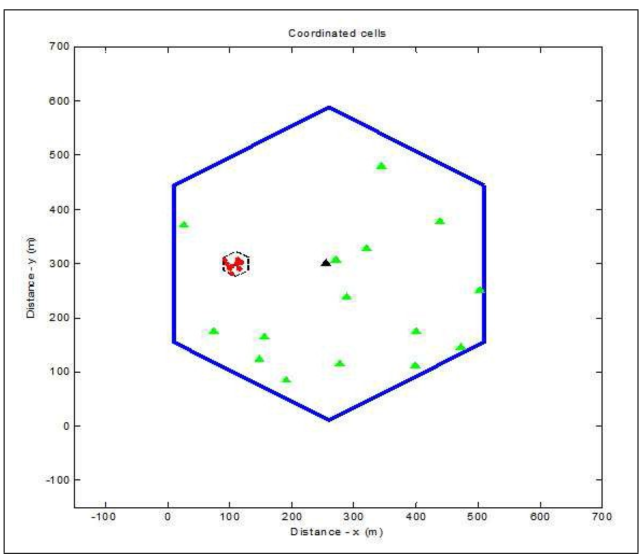

Fig. 3. Simulated UE distribution in multi-tier cells (macr0- and small-)

In our model, each multi-agent q-learning is implemented where each UE has a set of states and actions. Reward-based condition is implemented in the algorithm to enable the learning process. The details of our q-learning algorithm are as follows:

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Number of cells (N)	2 cells, one macro- and one small- cell
Cell radius	Macro- cell = 500m; small- cell = 40m
System bandwidth	3 MHz
Number of PRBs	15
BS antenna gain	Macro- = 14 dBi; Small- = 5dBi
UE antenna gain	0 dBi
Max UE transmit power	24 dBm
Noise power	-169 dBm/Hz

- **State** our model specify 2 different set of q-learning states, namely $\{PL_{edge}, PL_{center}\}$ based on uplink path loss information measured at UE. Due non-dynamic movement of UE, one-dimensional q-learning state is utilized. By identifying uplink path loss, we are able to test against a set of actions, where;
- **Action** comprises of a set of power level, l, specified from the maximum UE transmit power, a = $\{p_1, p_2, ..., p_l\}$
- **Reward** identify whether the selected action, at the specified state provide optimal SINR or not. In this case, two set of reward conditions are specified where:

$$r = \begin{cases} SINR, & \text{if } P_{action} > P_o + \beta L \\ 0, & \text{otherwise} \end{cases}$$

- **Exploration** our model employ nearly full exploration strategy (i.e. E = 0.9) at the beginning of the iteration to enable the agent to identify the best action. The exploration strategy is then reduced (i.e. E = 0.5) to allow the agent to explore and exploit at the same time. This is then reduced to the most minimal exploration (i.e. E = 0.05) and finally, full exploitation strategy is implemented.
- **Q-learning variables** the learning factors in our model is set to a fixed value where, learning rate, α = 0.5 and the reward discount, γ = 0.9.
- **Updated q-value** equation (3) is deployed to update the q-value based on selected state and action.

To briefly summarize the sequence process of our q-learning model, each agent (i.e. UE) starts by identifying the state based on the path loss calculated. From here, it selects an action, which is then used compute the reward. This is updated to produce q-value using equation (3) as mentioned earlier. The process continues until we reach convergence.

V. FINDINGS AND ANALYSIS

This section presents the findings of our study where our focus is to see the convergence of q-learning algorithm with respect to SINR. The SINR value will indicate the efficiency of applying DQL. We also compare our model with fractional power control elaborated in open loop power control subsection, based on equation (2). SINR for macro- and small- UE is computed using the following formula:

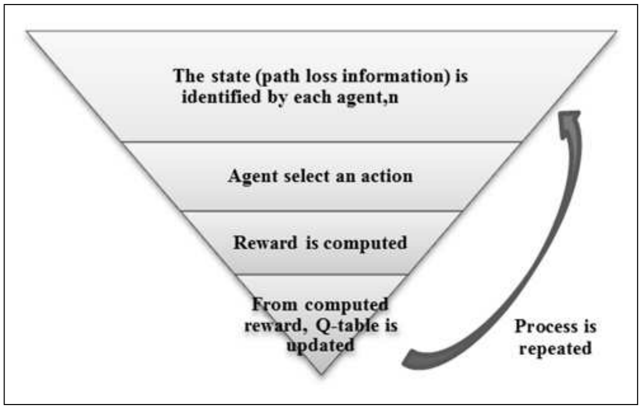

Fig. 4. Decentralized q-learning flow of process

$$SINR_{mbs} = \frac{P_{mue} * G_{mue} * G_{ant}}{I_{sue} + N} \quad (4)$$

$$SINR_{sbs} = \frac{P_{sue} * G_{sue} * G_{ant}}{I_{mue} + N} \quad (5)$$

Where P_{mue} and P_{sue} is the received power for macro- and small- UE respectively, G_{mue} and G_{sue} is macro- and small- UE antenna gain, G_{ant} is the BS antenna gain, I_{mue} and I_{sue} are co-tier interferences. We consider that each UE in the serving cell is interfered by one UE from the neighbouring cell. The interference power vary as the transmit power vary. This is due to q-learning implemented parallel for both macro- and femto- UE. The algorithm deploy epsilon-greedy policy where initial iteration, almost all actions are explored. This is gradually decreased until the q-value reached its optimality. We utilize simple path loss model (PL = 15.3 + 37.6 $\log_{10}(distance)$) where we disregard shadowing and penetration loss for the time being.

As mentioned earlier, the first part (40 per cent of the iterations) prompts the algorithm to explore principally all set of actions available. This is followed by partial (E = 0.5) and minimal (E = 0.05) exploration. Finally, full exploitation is applied. From here, we obtain convergence graph of accumulated SINR of all UEs. The result depicted by average SINR of all UE versus number of iterations is displayed in fig. 5. In the figure, the initial part has variation due to the exploration strategy employed. Meanwhile, there is increase in SINR on the second and the third parts of the exploration strategy. Finally, full exploitation can be seen at the last part of the iteration where the graph converges.

Further, fig. 6 demonstrates SINR comparison between our proposed method and fractional power control. In the figure, our proposed method is depicted by green block bar whilst fractional power control is represented by the red dotted bar. It can be seen that our proposed method has shown better performance with an increase factor of 3.5 as compared to conventional fractional power control.

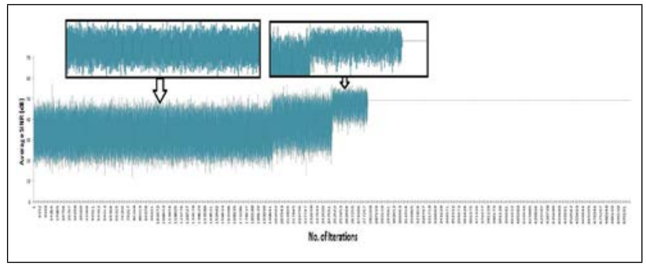

Fig. 5. Average SINR vs. Number of Iterations

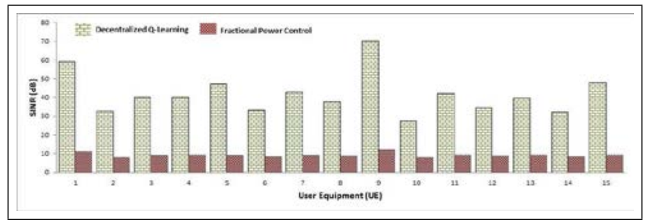

Fig. 6. SINR comparison between DQL and FPC

VI. CONCLUSION

It can be concluded from the results that our proposed method is able to optimize UE transmit power. This is crucial especially if UE is located at the edge of the cell. The implementation of DQL would enable UE to learn independently. This would allow UE to specify transmit power without the need to wait for transmission control from BS. In other words, it would reduce transmission establishment in uplink. Future works include the utilization of DQL in asymmetric transmissions to reduce multi-tier network load.

REFERENCES

[1] E. Dahlman, S. Parkvall and J. Skold, 4G LTE/LTE-Advanced for Mobile Broadband, Academic Press, 2013.

[2] P.Xia, V. Chandrasekhar and J. G. Andrews, ”Open vs Closed Access Femtocells in the Uplink,” IEEE Transactions on Wireless Communications, vol. 9, no. 12, pp. 3798-3809, December 2010.

[3] H. Zhang, N. Prasad, S. Rangrajan, S. Mekhail, S. Said and R. Arnott, ”Standard-compliant LTE and LTE-A Uplink Power Control,” 2012 IEEE International Conference in Communications, pp. 5275-5279, 2012.

[4] A. G-Serrano, L. Giupponi and G. Auer, ”Distributed Learning in Multiuser OFDMA Femtocell Networks,” IEEE 73rd Vehicular Technology Conference (VTC Spring), pp. 1-6, May 2011.

[5] A. G-Serrano and L. Giupponi, ”Distributed Q-Learning for Interference Control in OFDMA-based Femtocell Networks,” IEEE 71st Vehicular Technology Conference (VTC Spring), pp. 1-5, May 2010.

[6] H. Saad, A. Mohamed and T. ElBatt, ”A Cooperative Q-Learning Approach for Distributed Resource Allocation in Multi-User Femtocell Networks,” 2014 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1490-1495, April 2014.

[7] M. Simsek and A. Czylwik, ”Decentralized Q-Learning of LTE Femtocell for Interference Reduction in Heterogeneous Networks using Cooperation,” 2012 International ITG Workshop on Smart Antennas, pp. 86-91, March 2012.

[8] M. Bousif, C. Rosa, J. Wigard and R. Mullner,” Load Adaptive Power Control in LTE Uplink,” 2010 European Wireless Conference, pp. 288-293, April 2010.

[9] S. Deb and P. Monogioudis, ”Learning-based uplink Interference Management in 4G LTE Cellular Systems,” IEEE/ACM Transactions on Networking, vol. 23, no. 2, pp. 398-411, 2015.

[10] Y. Wang and M. Tao, ”Dynamic Uplink/Downlink Configuration Using Q-Learning in Femtocell Networks,” 2014 IEEE/CIC International Conference on Communications in China, pp.53-58, October 2014.
[11] S. Berger, B. Almeroth, V. Suryaprakash, P. Zanier, I. Viering, and G. Fettweis,”Dynamic Range-Aware Uplink Transmit Power Control in LTE Networks: Establishing an Operational Range for LTE’s Open-Loop Transmit Power Control Parameters (P_o, α),” IEEE Wireless Communications Letters, vol. 3, no. 5, pp. 521 - 524, October 2014.
[12] J. Gora, K. Pedersen, A. Szufarska, and S. Strzyz,”Cell-Specific Uplink Power Control for Heterogeneous Networks in LTE,” IEEE 72nd Vehicular Technology Conference Fall, pp. 1-5, Sept 2010.
[13] 3GPP-LTE ETSI TS 136 213 V8.6.0, ”Evolved Universal Terrestrial Radio Access (E-UTRA): Physical Layer Procedures,” April 2009.
[14] 3GPP, TSG RAN WG1 51 R1-074850, ”Uplink Power Control for E-UTRA Range and Representation of Po,” Ericsson, Nov. 2007.
[15] Y. Sun, R. P. Jover and X. Wang, ”Uplink Interference Mitigation for OFDMA Femtocell Networks,” IEEE Transactions on Wireless Communications, vol. 11, no. 2, pp. 614-624, February 2012.