Competitive Energy Allocation for Aerial Computation Offloading: A Colonel Blotto Game

Panagiotis Charatsaris*, Maria Diamanti*, Eirini Eleni Tsiropoulou†, and Symeon Papavassiliou*
{pancharatsaris@netmode.ntua.gr, mdiamanti@netmode.ntua.gr, eirini@unm.edu, papavass@mail.ntua.gr}
* School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece
† Dept. of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001655

***Abstract*—In this paper, we consider a competitive aerial computation offloading environment, where two edge resource operators provide computing services on a time-slot basis to multiple users, via Unmanned Aerial Vehicles (UAVs), each bearing a mounted edge server. The aim of each UAV is to selfishly maximize the difference between its personal and the opponent UAV's utility, by competitively allocating its energy resources to the different users in the system. The problem is formulated as a Generalized Colonel Blotto (GCB) game, where the UAVs allocate their resources across a number of battlefields, i.e., the users, as competing players, seeking to win the battlefield by increasing the difference of their in-between allocated resources and thus, experienced utility. The overall framework is complemented by a Reinforcement Learning (RL)-empowered algorithm to account for the energy efficient scheduling of the UAVs' overall available energy in the different time slots, where the GCB game is realized. The performance evaluation of the proposed framework is achieved via modeling and simulation. The obtained numerical results demonstrate the operation of the proposed GCB game, under different levels of competitiveness between the UAVs, and assess the effectiveness and efficiency of the proposed RL algorithm against different comparative scenarios.**

***Index Terms*—Unmanned Aerial Vehicles, Edge Computing, Colonel Blotto Game, Competition Modeling, Energy Efficiency.**

I. Introduction

By steadily consolidating the architecture concept of Multi-Access Edge Computing (MEC) in Next Generation (NextG) wireless networks, an increasing number of business entities will distribute and deploy MEC servers in different service areas, competing for the mobile users therein. Consequently, an environment of multifaceted competition will be created, where competition will not only originate from the user side, striving to share a common pool of computing resources, but also from the edge resource owners/operators that pursue their personal profit maximization. Economic models that capture the competitive interactions among business entities/operators and their impact at the end users have been widely explored and developed regarding different business sectors (e.g., Coca-Cola and Pepsi) [1], but are yet to be applied and tested in the field of wireless communications and computing.

This research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "1st Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: HFRI-FM17-2436).

In this paper, motivated by the research gap related to the problem of competitive resource allocation for aerial computation offloading in NextG MEC systems, we scrutinize the application of a Generalized Colonel Blotto (GCB) game [2] between two edge resource operators that compete with each other in selfishly maximizing the difference between their personal and the opponent's utility. Each operator deploys an Unmanned Aerial Vehicle (UAV), bearing a mounted MEC server, to provide computing services to the users in a geographical area, by properly allocating its available energy, and thus, computing power, to the multiple users per time slot, via the Colonel Blotto game. The framework is complemented by a Reinforcement Learning (RL) algorithm targeting the UAVs' overall available energy scheduling over a time horizon.

A. Related Work & Motivation

There exists a handful of research works that account for the business motives and the profit maximization seeking behavior of the different edge resource operators, during the computing resource allocation. In [3], a two-stage multi-server multi-user game is proposed, where the servers competitively determine the computing service's price, based on which the users' computation offloading decision follows via a non-cooperative game between them. A similar problem is addressed in [4], except that a third party (auctioneer) coordinates the introduced double auction mechanism, i.e., price setting and user-to-server matching. Following a diametrically opposite approach, the works in [5], [6] consider non-competitive environments and pursue the edge resource providers' profit maximization through the overall market's profit increase. The latter is achieved by designing multi-stage game-theoretic incentive mechanisms that result in mutually beneficial points for all the involved parties. Thus, the literature so far has mainly focused on the pure edge resource operators' profit maximization, by either competing or cooperating with each other, under different settings and formulations. To the best of the authors' knowledge, there does not exist works that consider the selfish and greedy behavior of an edge resource operator trying to maximize the difference between its personal and the opponents' utility, and strengthen its competitive position.

A two-player game-theoretic model that targets at exactly maximizing the difference between the two players' utilities is the Colonel Blotto game [7]. In this game, the players allocate their resources across a number of battlefields. The relative number of battlefields won, i.e., times that the allocated resources are more than the opponent's, determines each player's utility. Owing to its win-lose outcome, the Colonel Blotto game is one of the most appropriate candidate methods to solve problems related to jamming attacks in wireless networks. In [8], the problem of combating jamming in slow fading channels between a jammer and a defender, i.e., Access Point, across a number of battlefields, i.e., vulnerable users, is treated via the Colonel Blotto game, by determining the optimal power control strategy against all possible jammer power ranges. In [9], the Colonel Blotto game is utilized for secure wireless power transfer in aerial MEC networks, while other applications such as internet security, security in transportation systems, and rumor spread control in social networks are discussed and evaluated in [10]. The application of Colonel Blotto game in competitive resource management is studied in [11], where the dynamic spectrum allocation to users by different competing Network Service Providers (NSPs) is addressed. Nevertheless, the theory of the typical Colonel Blotto game in [7] that is followed by all aforementioned works, i.e., [8]–[11], results in mixed strategy solutions that are not practical for real-life applications. For this reason, generalizations of the Colonel Blotto game can be derived to conclude to pure Nash Equilibrium (NE) strategies, e.g., [2], but are yet to be applied and tested in real-life applications.

B. Contributions & Outline

In this paper, we address the research gap related to the competitive resource allocation for aerial computation offloading in NextG MEC networks, between two greedy edge resource operators, via the adoption of GCB game. The operators employ UAVs of fixed energy availability to provide computing services to the users of a geographic area over a time horizon. The main contributions of the paper are summarized as:

1) A competitive aerial computation offloading environment is considered, where two UAVs of different operators, selfishly allocate their available energy (and thus their computing power) to the multiple existing users in the area (battlefields) on a time-slot basis, with the aim to maximize the difference between their personal and the opponent's utility, via a GCB game.
2) The overall available energy of each UAV is scheduled in the different time slots of the considered time horizon via an RL algorithm, targeting at the maximization of the ratio between the corresponding UAV's achieved utility that stems from the GCB game and the actually consumed energy at the given time slot.
3) A unified framework is designed, where the UAVs' energy allocation to the users via the GCB game is iteratively optimized, while an RL-based energy scheduling procedure in the specific time slot is realized. Detailed numerical results demonstrate the operation and effectiveness of the proposed framework.

The remainder of this paper is organized as follows. Section II presents the system model and the overall framework's operation. Section III introduces the competitive energy and thus, computing power, resource allocation via the GCB game. Section IV describes the RL-based energy scheduling over the time horizon. Section V regards the overall framework's performance evaluation. Section VI concludes the paper.

II. System Model & Overall Framework

We consider a competitive aerial computation offloading environment, consisting of a set of users $\mathcal{N} = \{1, \dots, n, \dots, N\}$ and two competing serving UAVs i and j. The UAVs bear mounted MEC servers and offer computing services to the users for a time horizon $\mathcal{H}$, by hovering at their position. The time horizon $\mathcal{H}$ is equally divided in T time slots, and the user tasks' offloading and processing by the UAVs is performed on a per time slot basis. At each time slot t, each user n generates a number of computation tasks a_n^t according to some generalized process. Each user's n generated tasks a_n^t enter a queue in a First-In-First-Out (FIFO) mode at the node, awaiting for their offloading and processing by the UAVs i, j, along with a number of q_n^t computation tasks that potentially pre-exist in the queue from the previous time slots. The mean data size and data processing intensity of each user's n computation tasks are independently and identically distributed at each time slot t with mean values d_n [Bytes] and ϕ_n [CPU cycles/Byte], accordingly.

The UAVs i, j are characterized by an energy availability E_m^{tot} [J], $\forall m = \{i, j\}$, which should be scheduled for the user tasks' processing over the time horizon $\mathcal{H}$. At each time slot t, each UAV determines the specific energy level $E_m^t = E_m^{t,C} + E_m^{t,H}$ [J] to be scheduled for facilitating the processing energy consumption $E_m^{t,C}$ of the users' computation tasks existing in their queues and the UAV's hovering energy consumption $E_m^{t,H}$, where $0 \leq E_m^t \leq E_m^{tot} - \sum_{k=1}^{t-1} E_m^k$. Subsequently, the scheduled energy levels $E_m^{t,C}$ in the time slot t are shared among the N users, after a competitive allocation procedure takes place between the two UAVs in the form of a GCB game. The outcome of the GCB game is the energy levels $E_{n,m}^{t,C}$ allocated to each user n by each UAV $m = \{i, j\}$ at time slot t regarding the user tasks' processing. We denote the UAVs' computing capability as f_m [CPU cycles/sec], and assume that it is sufficient for processing in parallel the tasks. Given the determined energy level $E_{n,m}^{t,C}$, the number of the user's n computation tasks $w_{n,m}^t$ that are ultimately processed the UAV $m = \{i, j\}$ can be derived from the following formula:

$$w_{n,m}^t = \lfloor \frac{E_{n,m}^{t,C}}{\eta_m d_n \phi_n f_m} \rfloor, \tag{1}$$

where η_m indicates the mounted MEC server's at each UAV effective capacitance coefficient.

III. Generalized Colonel Blotto Game Formulation & Solution

In this section, we study and analyze the GCB game that is employed to formulate the interaction between the UAVs i, j during the competitive energy allocation to multiple serving users at a specific time slot, given the specific time slot's

t energy level $E_m^t, \forall m = \{i,j\}$. Then, we determine the solution of the game that is a pure NE strategy. Please note that in our proposed approach, the energy level E_m^t at each time slot t for each UAV is determined based on the RL algorithm, described in detail in Section IV.

A. Generalized Colonel Blotto Game Formulation

The interactions between the two UAVs $m = \{i,j\}$ regarding the competitive allocation of their energy level $E_m^{t,C}$ among the N users on a time slot t is formulated as a Generalized Colonel Blotto (GCB) game with N battlefields: $G = \{\mathcal{P}, \{Q_m^t\}_{m\in\mathcal{P}}, \{E_m^{t,C}\}_{m\in\mathcal{P}}, \mathcal{N}, \{v_n^t\}_{n=1}^N, \{U_m^t\}_{m\in\mathcal{P}}\}$. $\mathcal{P} = \{i,j\}$ is the set of players, i.e., the two competing UAVs and $Q_m^t = \{\mathbf{r}_m^t | \sum_{n=1}^N r_{n,m}^t \leq E_m^{t,C}, r_{n,m} \geq 0\}$ is each UAV's strategy space, where $r_{n,m}^t$ is the allocated energy of UAV $m = \{i,j\}$ to battlefield n and $\mathbf{r}_m^t \in Q_m^t$ is the vector of UAV's m overall energy allocations. The available energy of each UAV m for allocation to the battlefields is $E_m^{t,C}$, while $\mathcal{N} = \{1,\dots,n,\dots,N\}$ is the set of battlefields corresponding to the set of users existing in the area. The normalized value of each battlefield n is defined as

$$v_n^t = \frac{a_n^t + q_n^t}{\sum_{n=1}^N a_n^t + q_n^t}, \tag{2}$$

expressing the percentage of the number of computation tasks existing in user's n queue at time slot t, in relation to the total number of tasks in the system. The payoff $u_{n,m}^t$ that UAV m receives from battlefield n is defined as:

$$u_{n,m}^t(r_{n,m}^t, r_{n,-m}^t, k^t) = \frac{v_n^t}{1 + e^{-k^t(r_{n,m}^t - r_{n,-m}^t)}}, \tag{3}$$

where k^t is an approximation parameter that when tending to infinity yields the typical Colonel Blotto game defined in [7], while $r_{n,-m}^t$ is the energy allocated from the opponent of UAV m to battlefield n. The overall utility of UAV m at time slot t is $U_m^t = \sum_{n=1}^N u_{n,m}^t$, which is apparently dependent on the difference between its resources and the opponent's resources allocated across battlefields $n \in \mathcal{N}$, indicating each UAV's aim to maximize its personal utility, while reducing the opponent's. The UAVs compete by allocating different amounts of energy to each battlefield, i.e., user, at a given time slot. The UAV that invests more resources, i.e., energy, wins the battlefield. Nevertheless, even the player who loses experiences a utility, the value of which is determined by the selection of the approximation parameter k^t.

B. Pure-Strategy NE Solution

The pure-strategy NE solution of the considered GCB game can be derived by solving the following minimax problem:

$$V_m^t \triangleq 1 - V_{-m}^t \triangleq \min_{\mathbf{r}_{-m}^t \in Q_{-m}^t} \max_{\mathbf{r}_m^t \in Q_m^t} U_m^t(\mathbf{r}_m^t, \mathbf{r}_{-m}^t, k^t), \tag{4}$$

where V_m^t is the value of the GCB constant-sum game for the player m.

Proposition 1. *All strategies of the form $\sum_{n=1}^N r_{n,m}^t < E_m^{t,C}$ are dominated by the strategies $\sum_{n=1}^N r_{n,m}^t = E_m^{t,C}$, i.e., the UAV $m = \{i,j\}$ is better off using all of its available energy.*

Proof. The proposition is easily proved by contradicting the assumption that an allocation vector $\overline{\mathbf{r}}_m^t$, for which it holds $\sum_{n=1}^N \overline{r}_{n,m}^t = \overline{E}_m^{t,C} < E_m^{t,C}$, maximizes $U_m^t, \forall m$. □

To facilitate the subsequent analysis, we define the difference between the two UAVs' allocated energy on each battlefield $n \in \mathcal{N}$ at time slot t as $z_n^t = r_{n,i}^t - r_{n,j}^t$, and the difference between their available energy resources as $D^t = E_i^{t,C} - E_j^{t,C}$, thus, $\sum_{n=1}^N z_n^t = D^t$.

Without loss of generality, we assume that UAV i has more available energy than j at time slot t, i.e., $E_i^{t,C} > E_j^{t,C}$, the values of the battlefields are sorted in ascending manner, i.e., $v_1^t \geq \cdots \geq v_N^t$, with one of these inequalities being strict to avoid the case of all battlefields being identical, and $\frac{E_i^{t,C}}{N} < E_j^{t,C}$ to eliminate the case that UAV i trivially wins the game. The latter yields $0 < D^t < (N-1)E_j^{t,C}$.

Theorem 1. *A local maximum $\mathbf{z}^{t*} = [z_1^{t*}, \dots, z_N^{t*}]$ of $U_i^t(\mathbf{z}^t, k^t)$ is a solution to the following equation:*

$$\sum_{n=1}^{N-1} z_n^{t*} + z_N^t = D^t, \tag{5}$$

where

$$z_n^{t*} = \frac{1}{k^t} \ln\Big[\frac{v_n^t(1+e^{-k^t z_N^t})^2 - 2v_N^t e^{-k^t z_N^t}}{2v_n^t e^{-k^t z_N^t}} + \frac{(1+e^{-k^t z_N^t})\sqrt{v_n^t(v_n^t(1+e^{-k^t z_N^t})^2 - 4v_N^t e^{-k^t z_N^t})}}{2v_N^t e^{-k^t z_N^t}}\Big], \tag{6}$$

for $n \in \mathcal{N} - \{N\}$ and $z_N^t > 0$.

Proof. If $\mathbf{z}^{t*}$ is a local mazimizer and U_i^t is continuously differentiable, then $\nabla U_i^t(\mathbf{z}^{t*}, k^t) = 0$. Considering that $z_N^t = D^t - \sum_{n=1}^{N-1} z_n^t$ and after some manipulations, we conclude that Eq. (5) is the only possible local maximum of $U_i^t(\mathbf{z}^t, k^t)$. Also, the Hessian matrix $\nabla^2 U_i^t(\mathbf{z}^{t*}, k^t)$ must be negative definite. Taking all the different cases it is proved that this is true only if $z_N^t > 0$. □

We define the function $f_k(z_N^t) = \sum_{n=1}^{N-1} z_n^{t*} + z_N^t = D^t$, for which the following proposition holds true.

Proposition 2. *The function $f_k(z_N^t)$ is strictly convex, whose minimum occurs at a negative z_N^t, i.e., $z_N^{t,min} < 0$.*

Proof. The first part of the proposition is proved by calculating the value of the second order derivative of $f_k(z_N^t)$ with respect to z_N^t, while the second part is proved by observing that $z_N^{t,min} < 0$ for the equation $\frac{df}{dz_N^t}|_{z_N^t = z_N^{t,min}} = 0$ to solve. □

We assume that $\underline{D}^t = f_k(z_N^{t,min})$ is the minimum value of $f_k(z_N^t)$, which is unique since $f_k(z_N^t)$ is strictly convex.

Corollary 1. $f_k(z_N^t) = D^t$ *has: a) two solutions if $D^t > \underline{D}^t$, b) one solution if $D^t = \underline{D}^t$, or no solution if $D^t < \underline{D}^t$.*

Thus, for $D^t \geq \underline{D}^t$ there exists one or two possible local maxima for $U_i^t(\mathbf{z}^t, k^t)$ that must lead to a positive solution for

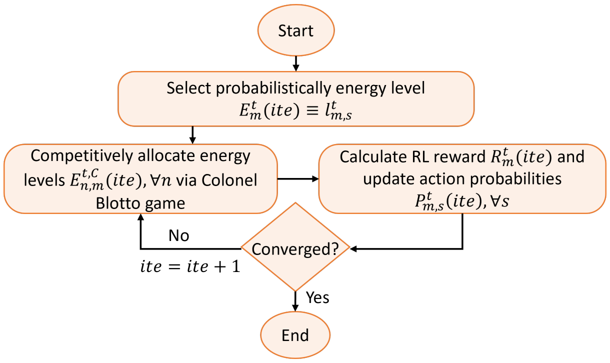


Fig. 1: Unified framework overview.

z^{t*}_N (from Theorem 1) and satisfy $r^t_{n,i} \leq E^{t,C}_i, r^t_{n,j} \leq E^{t,C}_j \Rightarrow -E^{t,C}_j \leq z^t_n = r^t_{n,i} - r^t_{n,j} \leq E^{t,C}_i$. Thus, we conclude that for $n \in \mathcal{N}$ any solution must satisfy $0 < z^{t*}_n \leq E^{t,C}_i$.

Theorem 2. *For* $D^t k^t \geq \sum_{n=1}^{N-1} \ln \frac{2v^t_n - v^t_N + 2\sqrt{v^t_n(v^t_n - v^t_N)}}{v^t_N}$ *and* $1 + e^{\frac{k^t D^t}{N}} \leq \sqrt{v^t_N} e^{\frac{k^t D^t}{2N}} e^{k^t D^t \frac{N}{2(N-1)}}$, $U^t_i(\mathbf{z}^t, k)$ *obtains a unique maximum.*

Proof. It can be easily proved that $f_k(z^t_N)$ is always greater than its asymptotes, lying in the upper subspace of their intersection. Assume that $\underline{z}^{t*}_N$ and $\overline{z}^{t*}_N$ are the lower and higher solutions of $f_k(z^t_N) = D^t$ and, hence, $\overline{z}^t_N > \overline{z}^{t*}_N \geq z^{t,min}_N \geq \underline{z}^{t*}_N > \underline{z}^t_N$, where $\underline{z}^t_N, \overline{z}^t_N$ are the intersection points of line $y = D^t$ and the right and left asymptotes, respectively. From Proposition 2, we have that $z^{t,min}_N < 0$ and for this reason $\underline{z}^{t*}_N$ cannot be an acceptable maximum. Studying $\overline{z}^{t*}_N$, we can prove that $\overline{z}^t_N < E^{t,C}_i \Rightarrow \overline{z}^{t*}_N < E^{t,C}_i$. But $\overline{z}^{t*}_N$ must be, also, positive $\iff$ the first inequality of this Theorem holds true. Finally, the resulting z^{t*}_n must also be lower than $E^{t,C}_i$. From this condition, we conclude to $1 + e^{\frac{k^t D^t}{N}} \leq \sqrt{v^t_N} e^{\frac{k^t D^t}{2N}} e^{k^t D^t \frac{N}{2(N-1)}}$. □

Theorem 3. *The unique local maxima* $\mathbf{z}^{t*} = [z^{t*}_1, \ldots, z^{t*}_N]$, *where* $z^{t*}_N = \overline{z}^{t*}_N$ *and* z^{t*}_n *arises from Theorem's 1 relation for* $n \in \mathcal{N} - \{N\}$, *is a unique global maximum of* $U^t_i(\mathbf{z}^t, k^t)$ *for* $D^t k^t \geq \sum_{n=1}^{N-1} \ln \frac{2v^t_n - v^t_N + 2\sqrt{v^t_n(v^t_n - v^t_N)}}{v^t_N}$ *and* $1 + e^{\frac{k^t D^t}{N}} \leq \sqrt{v^t_N} e^{\frac{k^t D^t}{2N}} e^{k^t D^t \frac{N}{2(N-1)}}$.

Proof. This proof follows same logic as that of Theorem 1. □

Based on the preceding analysis, Theorem 4 concludes the pure-strategy NE of the proposed GCB game.

Theorem 4. *For* $D^t k^t \geq \sum_{n=1}^{N-1} \ln \frac{2v^t_n - v^t_N + 2\sqrt{v^t_n(v^t_n - v^t_N)}}{v^t_N}$ *and* $1 + e^{\frac{k^t D^t}{N}} \leq \sqrt{v^t_N} e^{\frac{k^t D^t}{2N}} e^{k^t D^t \frac{N}{2(N-1)}}$ *the pure-strategy NE for the GCB game comprises the following allocation vectors:*

$$\mathbf{r}^{t*}_j = [r^{t*}_{1,j}, \ldots, r^{t*}_{N,j}], \tag{7}$$
$$\mathbf{r}^{t*}_i = \mathbf{r}^{t*}_j + [z^{t*}_1, \ldots, z^{t*}_N], \tag{8}$$

where $z^{t*}_N = \overline{z}^{t*}_N$ *(positive solution of* $f_k(z^t_N) = D^t$*), and* z^{t*}_n, *for* $n \in \mathcal{N} - \{N\}$ *derives from Eq. (6) and* $r^{t*}_{1,j} + \cdots + r^{t*}_{N,j} = E^{t,C}_j$.

Proof. Based on the minimax problem in Eq. (4), maximizing $U^t_i(\mathbf{r}^t_i, \mathbf{r}^t_j, k^t)$ over $\mathbf{r}^t_i$ is similar to maximizing $U^t_i(\mathbf{z}^t, k^t)$ over $\mathbf{z}^t = \mathbf{r}^t_i - \mathbf{r}^t_j$. Theorem 3 proved that the utility function has a unique maxima. Therefore, $\mathbf{r}^{t*}_i = \mathbf{r}^{t*}_j + \mathbf{z}^{t*}$. By substituting $\mathbf{r}^{t*}_i$, the utility function has a constant value. Consequently, $\mathbf{r}^{t*}_j$ consists of all the strategies in Q^t_j and, hence, the pure-strategy NE of the GCB is as defined in (7)-(8). □

By obtaining each UAV's $m = \{i, j\}$ optimal strategies $\mathbf{r}^{t*}_m$, i.e., allocated energy to each battlefield/user from the scheduled energy $E^{t,C}_m$ in slot t, then, the number of each user's computation tasks to be processed derives from Eq. (1).

IV. RL-Empowered Energy Scheduling over Time

In this section, an RL mechanism is introduced to enable each UAV's $m = \{i, j\}$ energy scheduling E^t_m in the time slots $t \in \mathcal{H}$ in an autonomous and distributed manner. Please be reminded that the estimated energy levels E^t_m, and hence, $E^{t,C}_m$, are used as input to the GCB game for allocation to the users. In particular, we employ a gradient ascent RL algorithm and model the UAVs as stochastic learning automata. The RL algorithm comprises a number of iterations, denoted as ite, during which each UAV probabilistically selects a specific energy level to be scheduled in time slot t from a finite set of S energy levels/actions $\mathcal{L}^t_m = \{l^t_{m,1}, \ldots, l^t_{m,s}, \ldots, l^t_{m,S}\}$. Apparently, the different available energy levels/actions $l^t_{m,s} \in \mathcal{L}^t_m, \forall s \in S$ are functions of the corresponding UAV's available energy at time slot t, i.e., $E^{t,avail}_m = E^{tot}_m - \sum_{k=1}^{t-1} E^k_m$. The selection of a specific energy level $E^t_m \equiv l^t_{m,s} \in \mathcal{L}_m$ yields a corresponding reward $R^t_{m,s}$ to UAV m of the perceived satisfaction from scheduling this energy level in the time slot t, defined as

$$R^t_{m,s}(ite) = \frac{\sum_{i=1}^{ite} U^t_m(i)}{cE^t_m(ite)}. \tag{9}$$

where c $[J^{-1}]$ is the unit cost of consumed energy. Eq. (9) captures the tradeoff between the experienced utility stemming from the GCB game over the past iterations and the invested energy level, aiming at the system's energy efficient operation.

The probability of selecting a specific energy level $E^t_m \equiv l^t_{m,s} \in \mathcal{L}^t_m$ at the next RL iteration $(ite + 1)$ is updated based on the Linear Reward Inaction (LRI) gradient ascent algorithm that applies the following probability update rules:

$$P^t_{m,s}(ite+1) = P^t_{m,s}(ite) + b \cdot R^t_{m,s}(ite) \cdot (1 - P^t_{m,s}(ite)),$$
$$\text{if } l^t_{m,s}(ite+1) = l^t_{m,s}(ite). \tag{10}$$

$$P^t_{m,s}(ite+1) = P^t_{m,s}(ite) - b \cdot R^t_{m,s}(ite) \cdot P^t_{m,s}(ite),$$
$$\text{if } l^t_{m,s}(ite+1) \neq l^t_{m,s}(ite). \tag{11}$$

Note that $0 < b < 1$ is the learning rate parameter that controls the RL algorithm's tradeoff between exploration and exploitation [12]. The UAVs i, j iteratively select energy levels to schedule in a time slot t, while they learn the most energy efficient one. The overall RL algorithm converges, when $P^t_{m,s}(ite) \to 1, \forall m = \{i, j\}$. An overview of the unified framework, highlighting the operation of both the GCB game and the RL mechanism, is presented in Fig. 1.

V. Evaluation & Results

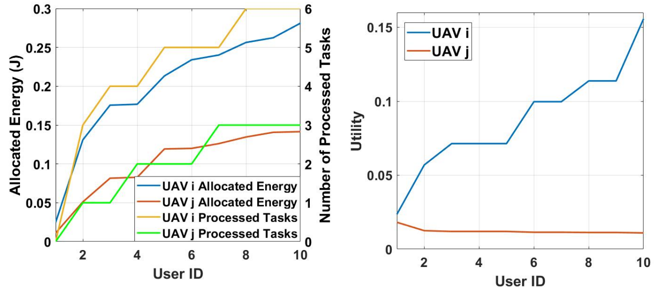
In this section, we evaluate the performance of the proposed framework, via modeling and simulation. In our simulation, we consider two UAVs i, j providing computing services to a set of $N = 10$ users over a time horizon of $\mathcal{H} = 10$ sec, which is equally divided in $T = 10$ time slots. At each time slot, each user generates a number of tasks according to a Poisson process with rate 6 tasks/sec, whose mean data size and processing intensity are set as $d = 1.7 \times 10^9$ Bytes and $\phi = 6.53 \times 10^3$ CPU cycles/Byte. The UAV-mounted MEC server's characteristics are set as $f_i = f_j = 8 \times 10^{12}$ CPU cycles/sec and $\eta_i = \eta_j = 8 \times 10^{-28}$. The UAVs are equipped with different initial energy availability, namely $E_i^{tot} = 20$ J and $E_j^{tot} = 15$ J, while their hovering energy consumption is assumed equal to $E_i^{t,H} = E_j^{t,H} = 10^{-2}$ J. Throughout our experiments, we consider the value of the parameter k equal to the lower bound of its feasible set (Theorem 2), unless otherwise explicitly stated. Regarding the RL algorithm, we consider three possible energy levels, corresponding to 50%, 100% and 150% of the share that results if the available energy is equally divided by the remaining time slots, while the learning rate of the RL algorithm is set as $b = 0.65$ and the unit cost of consumed energy as $c = 1\ J^{-1}$.

A. Evaluation of Generalized Colonel Blotto Game

First, we study the pure operation of the GCB game under a given time slot, and subsequently, drop the superscript t for notation simplification purposes. The energy levels chosen by the UAVs via the RL algorithm are $E_i = 2$ J and $E_j = 1$ J, with $E_i > E_j$. Fig. 2a depicts the allocated energy levels (left vertical axis) and number of processed computation tasks (right vertical axis) by the UAVs i, j with respect to each user, denoted by the user ID index in the horizontal axis. It should be noted that the user ID index indicates the users sorted increasing order regarding their corresponding values of battlefields, i.e., $v_1 \leq \cdots \leq v_N$. UAV i, having more available energy than UAV j, allocates more energy and, thus, processes more tasks compared to UAV j, given that all user tasks in the system bear the same mean data size and processing intensity. Additionally, it can be easily observed that as the user ID increases, denoting users of higher computation burden and battlefields of higher values, then both the allocated energy and number of processed tasks increase for UAVs i, j, respectively.

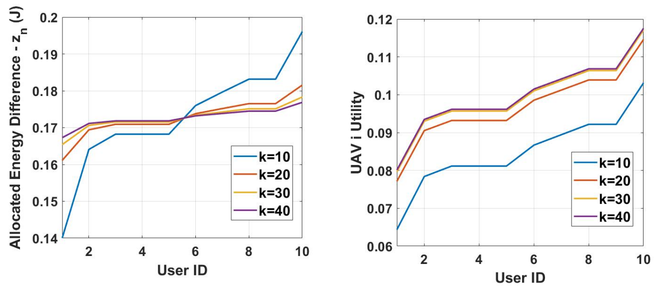
Fig. 2b illustrates the utilities obtained by the two UAVs i, j, when investing their energy resources across each battlefield, i.e., user. Apparently, the utility values of the two UAVs follow the exactly opposite trend as the user ID gets higher, with the utility of UAV i being constantly higher than the one of UAV j considering all battlefields, i.e., users. Also, it is observed that although the rate of increase in UAV's i utility increases significantly as the user ID gets higher, the rate of decrease of the UAV's j utility decreases, which is implied by the fact that GCB game is a constant-sum game.

Subsequently, in Fig. 3, we study the operation of the GCB game under different values of the approximation parameter



(a) (b)

Fig. 2: GCB game operation: (a) allocated energy and number of processed tasks, (b) utility of each UAV.



(a) (b)

Fig. 3: GCB game operation for different values of k: (a) allocated energy difference between UAVs, (b) UAV's i utility.

$k = \{10, 20, 30, 40\}$, in order to highlight the different competition levels that the GCB game allows between the competing UAVs. Specifically, Fig. 3b illustrates the allocated energy difference z_n across all battlefields $n \in \mathcal{N}$, which exhibits an increasing behavior as the user ID increases, considering all four cases of different values of k. For high user ID indices, lower values of the parameter k result in higher allocated energy differences to the users, whereas the opposite holds true for low user ID indices. The latter behavior is justified by the constant-sum game property of the GCB game. Last, in Fig. 3b, the utility of UAV i across the different battlefields (i.e. users) is depicted for different values of k. As k increases and the game tends to the typical Colonel Blotto, the UAV i that has more initial resources achieves a higher utility value in all battlefields/users, while the strictly increasing behavior of the UAV's utility as the user ID increases is maintained. The exact opposite observations and results hold for the UAV's j utility, which, however, are omitted due to space limitations. Thus, higher values of k aggravate the game's win-lose outcome and the competition between the UAVs.

B. Evaluation of RL-empowered energy scheduling algorithm

In the following, we study the performance of the RL-enabled energy scheduling algorithm over the time horizon. Fig. 3 illustrates the convergence behavior of the action probabilities for UAV i (Fig. 3a) and UAV j (Fig. 3b), revealing that almost after 10 RL iterations (corresponding to 2.22 secs in real execution time) the convergence is reached and each UAV's energy level to be scheduled in the corresponding time

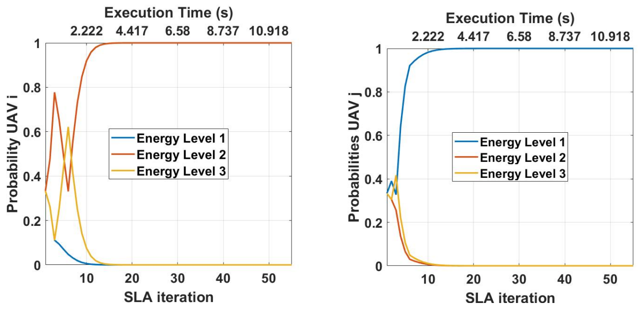


Fig. 4: RL algorithm convergence analysis: (a) UAV's i action probabilities, (b) UAV's j action probabilities.

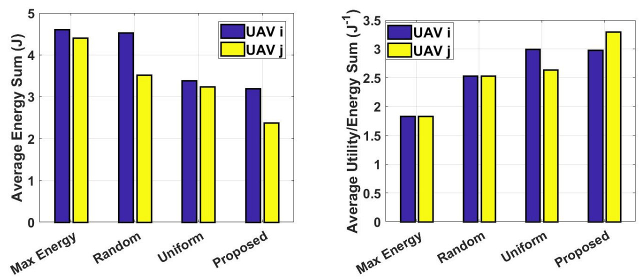


Fig. 5: RL algorithm comparative analysis: UAVs' (a) average energy sum, (b) ratio of average GCB utility to energy sum.

slot is determined. It is important to mention that UAV i, which initially owned more energy, ended up with higher energy level scheduled in the specific time slot, without however wasting its overall available energy. The latter behavior is motivated by the designed RL reward function in Eq. (9), which intelligently controls the tradeoff between "winning" the GCB game against the opponent UAV i and saving energy.

The performance of the proposed RL algorithm is compared against three different approaches, namely: a) the "Max Energy", where the higher energy level from the considered action set of the RL algorithm is chosen at each time slot, until the UAVs' energy is fully exhausted, b) the "Uniform" distribution of the available energy at the remaining time slots, and c) the "Random" selection of the energy level from the considered action set of the RL algorithm. In Fig. 5a, the sums of the average consumed energy over the considered time horizon are presented for both UAVs, verifying that the proposed RL-empowered algorithm concludes to a more energy efficient operation point. Complementary to Fig. 5a, Fig. 5b depicts each UAV's achieved ratio of the average utility obtained by the GCB game to the total consumed energy, over the time horizon. Once again, the numerical results confirm that the proposed RL algorithm manages to achieve the best tradeoff between the satisfaction derived from the battlefields won and the actually consumed energy.

VI. Conclusion & Future Work

In this paper, the competitive energy allocation problem based on the Generalized Colonel Blotto (GCB) game is studied, to support the aerial computation offloading of the users to two competing UAVs that bear mounted MEC servers. The UAVs, being characterized by some initial level of energy availability, aims to maximize the difference between their personal and the opponent UAV's utility, by competitively allocating energy resources to the users that play the role of the battlefields. The allocated energy to the users by the UAVs is translated into computing power to facilitate their computing needs for a given time slot. The overall framework is complemented by an RL algorithm to support the energy efficient scheduling of the UAVs' overall available energy in different time slots, within which the GCB takes place. The numerical results demonstrate the operational characteristics of the GCB game and the different levels of competitiveness allowed between the UAVs, as well as the effectiveness and efficiency of the proposed RL algorithm in terms of the actually consumed energy by each UAV. Part of our current and future work focuses on elaborating on the features of the GCB game regarding the competitiveness cultivated between the players and the exploration of a wider range of competitive resource allocation problems in the field of wireless communications.

References

[1] M. Chalikias and M. Skordoulis, "Implementation of fw lanchester's combat model in a supply chain in duopoly: the case of coca-cola and pepsi in greece," *Oper. Research*, vol. 17, no. 3, pp. 737–745, 2017.

[2] A. Ferdowsi, A. Sanjab, W. Saad, and T. Basar, "Generalized colonel blotto game," in *2018 Annual American Control Conference (ACC)*, 2018, pp. 5744–5749.

[3] Z. Chen, Q. Ma, L. Gao, and X. Chen, "Edgeconomics: Price competition and selfish computation offloading in multi-server edge computing networks," in *19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*, 2021, pp. 1–8.

[4] Y.-Y. Shih, A.-C. Pang, T. He, and T.-C. Chiu, "A multi-market trading framework for low-latency service provision at the edge of networks," *IEEE Trans. on Services Computing*, pp. 1–1, 2021.

[5] Q. Wang, S. Guo, J. Liu, C. Pan, and L. Yang, "Profit maximization incentive mechanism for resource providers in mobile edge computing," *IEEE Trans. on Services Computing*, vol. 15, no. 1, pp. 138–149, 2022.

[6] D. Kim, H. Lee, H. Song, N. Choi, and Y. Yi, "Economics of fog computing: Interplay among infrastructure and service providers, users, and edge resource owners," *IEEE Transactions on Mobile Computing*, vol. 19, no. 11, pp. 2609–2622, 2020.

[7] B. Roberson, "The colonel blotto game," *Economic Theory*, vol. 29, no. 1, pp. 1–24, 2006.

[8] A. Gouissem, K. Abualsaud, E. Yaacoub, T. Khattab, and M. Guizani, "Game theory for anti-jamming strategy in multichannel slow fading iot networks," *IEEE Internet of Things Journal*, vol. 8, no. 23, pp. 16 880–16 893, 2021.

[9] Y. Wang, L. Zhang, M. Min, C. Guo, V. Sharma, and Z. Han, "Privacy-aware wireless power transfer for aerial computation offloading via colonel blotto game," in *IEEE Globecom Workshops*, 2020, pp. 1–6.

[10] S. Guan, J. Wang, H. Yao, C. Jiang, Z. Han, and Y. Ren, "Colonel blotto games in network systems: Models, strategies, and applications," *IEEE Trans. on Net. Sci. and Eng.*, vol. 7, no. 2, pp. 637–649, 2020.

[11] M. Hajimirsadeghi, G. Sridharan, W. Saad, and N. B. Mandayam, "Inter-network dynamic spectrum allocation via a colonel blotto game," in *Annual Conf. on Inform. Science and Syst. (CISS)*, 2016, pp. 252–257.

[12] Y. Xu, J. Wang, and Q. Wu, "Distributed learning of equilibria with incomplete, dynamic, and uncertain information in wireless communication networks," *Game Theory Framework Applied to Wireless Comm. Net., IGI Global*, pp. 63–86, 2016.