

Distributed Offloading in Multi-Access Edge Computing Systems: A Mean-Field Perspective

Shubham Aggarwal^{ID}, *Graduate Student Member, IEEE*, Muhammad Aneeq uz Zaman^{ID},
Melih Bastopcu^{ID}, *Member, IEEE*, Sennur Ulukus^{ID}, *Fellow, IEEE*, and Tamer Başar^{ID}, *Life Fellow, IEEE*

Abstract—With the widespread adoption of internet-of-things (IoT) devices capable of supporting numerous intelligent applications, the demand for computational power has surged dramatically. Multi-access edge computing (MEC) technology is a promising solution to assist the often power-constrained IoT devices by providing additional computing resources for time-sensitive tasks. In this paper, we consider the problem of optimal task offloading in MEC systems with due consideration of the timeliness and scalability issues under two scenarios of equitable and priority access to the edge server (ES). In the first scenario, we consider a MEC system consisting of N devices assisted by one ES, where the devices can split task execution between a local processor and the ES, with *equitable access* to the ES. In the second scenario, we consider a MEC system consisting of one primary user, N secondary users and one ES. The primary user has *priority access* to the ES while the secondary users have *equitable access* to the ES amongst themselves. In both scenarios, due to the power consumption associated with utilizing the local resource and task offloading, the devices must optimize their actions. Additionally, since the ES is a shared resource, other users' offloading activity serves to increase latency incurred by each user. We thus model both scenarios using a *large user non-cooperative game framework*. However, the presence of a large number of users makes it nearly impossible to compute the equilibrium offloading policies for each user, which would require a significant communication overhead to exchange information with each other. Thus, to alleviate such scalability issues, we invoke the paradigm of mean-field games (MFGs) to design completely distributed low complexity algorithms for the computation of approximate Nash equilibrium policies for each user based on only their local information. Further, by leveraging the novel age of information (AoI) metric, we study the trade-offs between increasing

information freshness and reducing power consumption for each user. Using numerical evaluations, we show that our approach can recover the offloading trends displayed under centralized solutions, and provide additional insights into the results obtained.

Index Terms—Multi-access edge computing, mean-field games, age of information.

I. INTRODUCTION

THE multi-access edge computing (MEC) technology has recently garnered significant attention as a potential solution to enhance computing capabilities in power-limited internet-of-things (IoT) devices [2]. The MEC architecture capitalizes on the benefits of wireless communication and mobile computing paradigms, thereby enabling the offloading of task execution to the network edge. This is in contrast with the traditional cloud computing technology, which is (1) located geographically far from the end-users, and (2) have limited number of computing platforms. The MEC technology, on the other hand, brings computing capability to the edge of the network, which is much closer to the consumer, increasing the number of computing stations, each serving a smaller number of users within a certain spatial area. This approach is expected to be crucial in time-sensitive applications such as vehicle positioning in autonomous driving, task assignment in warehouses, and remote surgery systems [2], [3], [4].

In this work, we aim to (1) increase the time responsiveness of task execution in MEC systems by employing the novel age of information (AoI) metric [5] to ensure the timeliness of task execution in time-sensitive applications, and (2) promote scalability of the paradigm by constructing completely distributed policy design methods using mean-field games (MFGs). Specifically, we first consider a MEC system with N devices assisted by an edge server (ES), which can be equitably accessed by each device.¹ Such a prototypical MEC system is shown in Fig. 1. The devices aim to optimally utilize the onboard processor power and the computation facility provided by the ES. Since the ES is a shared resource, the offloading policy of each user is affected by those of the other users. Thus, we aim to find Nash equilibrium policies for each user to balance between the power consumed at each device and the timeliness incurred by it. To capture the latter, we invoke the novel AoI metric to set up each device's problem as a multi-objective optimization problem

Received 26 June 2025; revised 23 October 2025; accepted 1 December 2025. Date of publication 11 December 2025; date of current version 7 May 2026. The work of Shubham Aggarwal, Muhammad Aneeq uz Zaman, Melih Bastopcu, and Tamer Başar was supported in part by ARO MURI under Grant AG285 and in part by AFOSR under Grant FA9550-24-1-0152. The work of Melih Bastopcu was supported by BIL2: BILKENT University-TUBITAK BILGEM Consultancy Call for Research EDGE-4-IoT and TUBITAK 2232-B under Project 124C533. Recommended for acceptance by J. Choi. (*Corresponding author: Shubham Aggarwal.*)

Shubham Aggarwal is with the Coordinated Science Laboratory (CSL) and the Department of Mechanical Science and Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801 USA (e-mail: sa57@illinois.edu).

Muhammad Aneeq uz Zaman is with Analog Devices Inc., Wilmington, MA 01887 USA (e-mail: mazaman2@illinois.edu).

Melih Bastopcu is with the Department of Electrical and Electronics Engineering, Bilkent University, 06800 Ankara, Türkiye (e-mail: bastopcu@bilkent.edu.tr).

Sennur Ulukus is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: ulukus@umd.edu).

Tamer Başar is with the CSL and the Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801 USA (e-mail: basar1@illinois.edu).

Digital Object Identifier 10.1109/TMC.2025.3642928

¹ We use the words user and device interchangeably in the sequel.

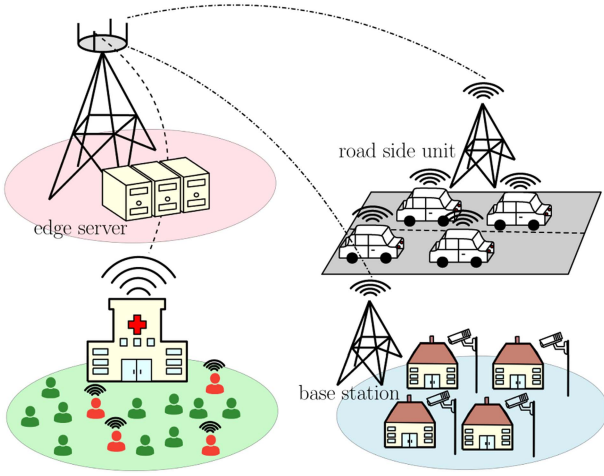


Fig. 1. The figure shows a prototypical MEC system consisting of an edge server and intelligent applications, such as connected autonomy, medical internet-of-things and surveillance that simultaneously utilize edge server for timely computation.

of balancing power consumption and the average AoI incurred by the offloaded packets. Additionally, to alleviate the issue of computational tractability of equilibrium policy computation posed by a large number of devices, we leverage the framework of (MFGs) [6], [7], [8], [9], [10] to compute distributed equilibrium policies based on each user's local information only. This allows us to decrease the communication overhead incurred as a result of information exchange required between users to obtain a Nash equilibrium solution. Consequently, our algorithm can scale to a large user population without additional costs. In addition, motivated by scenarios in *cognitive radio network* (CRN) technology [11], we also extend the aforementioned MEC architecture (with equitable access) to the one constituting a primary user, who has priority access to the ES while all other N users have equitable (but secondary) access to the ES via the primary user's transmitter. We model this problem, again, within a non-cooperative game framework, and consequently invoke the paradigm of major-minor mean-field games (MM-MFGs) [12] to alleviate the computational tractability issue posed by the large number of users. We provide distributed algorithms to compute (local) Nash equilibrium policies for the cases with and without the primary user. Finally, we corroborate the theoretical findings using extensive numerical evaluations.

A. Related Works

The subject of computation offloading in MEC systems has received wide attention in the past decade or so. One line of research in the area has focused on dedicated resource allocation of a portion of the total bandwidth to the involved users [13], [14], [15]. Specifically, [13] formulated the offloading problem as a power consumption minimization problem with constraints on the task buffer stability; with extensions provided in [15], [16] to incorporate wireless power transfer to the IoT devices. Consequently, the authors provided online algorithms that determine the local execution and computation offloading policy based on the Lyapunov penalty-plus-drift-based optimization

technique [17] and study the trade-offs between the power consumed and delay incurred, as a function of a control parameter. While the above works do not take into account timeliness considerations serving time-critical applications, a few recent works [4], [18], [19] have focused on timeliness within the realm of the MEC resource allocation. Specifically, [18] considered a single-source single-destination MEC system for timely status updating; [19] leveraged energy harvesting in addition to the MEC to support computing capabilities of the IoT devices; and [4] jointly assessed the impact of stochastic arrivals, scheduling policy, and unreliable channel conditions on the expected sum of AoI in MEC assisted IoT networks. We further refer the interested readers to surveys [20], [21] for additional details.

In contrast, to solve the high time-complexity issues faced by the central resource allocation techniques [22], relatively recent works use the game-theoretic framework [23], [24], [25], [26] for designing Nash or Stackelberg equilibrium offloading policies, which take into account the self-interested nature of the involved users. Specifically, the work [23] develops a cooperative-competitive game based optimization problem to compute offloading policies for monitoring health in internet-of-medical-things applications; [24] considers offloading decisions to minimize the delay subject to an offloading cost and design incentive schemes to utilize parked vehicles as edge computers within the MEC system; [25] considers a non-sequential offloading strategy based on a non-cooperative game approach which is effective in reducing latency compared to the traditional way of transmitting, planning, forwarding, and executing sequentially; and [26] formulates an offloading problem constituting a cloud server and multiple edge servers and designs incentive mechanisms for utility maximization of all the servers in a Stackelberg game setting.

Computation of Nash equilibria in the above settings, however, poses a challenge where each user is required to have the knowledge of the policy structure of the other users, since their cost functions are coupled by the presence of a shared resource. This requires not only an additional storage mechanism, but also a significant communication overhead for information exchange, especially for a large user population. This makes it difficult to scale the MEC paradigm to large user settings—a feature which is core to the MEC paradigm. To alleviate this issue of tractable policy computation, the framework of MFGs was proposed in the control systems literature, independently and concurrently in the works [7], [27] and [8] which, under suitable assumptions of homogeneity and anonymity among users, allows the computation of approximate distributed Nash equilibrium solutions without the necessity of information exchange between users. Using motivations from statistical physics, it exploits a limiting system (called a mean-field (MF) system) with $N = \infty$ to compute consistent equilibrium solutions for a “generic” user who is representative of the entire population. Such a framework has been widely applied in various large-user domains such as epidemiology, power systems, semantic control systems, wireless networks, satellite communications, finance [10], [28], [29], [30], [31], to name a few. Additionally, the setup of MFGs has also been extended to cases constituting the presence of an “influential” player in addition to the large

population of “minor” players and is termed as major-minor MFGs [12]. Inspired by financial markets and banking systems, such a formulation allows the study of the non-vanishing effects of players which hold major portfolios within the population.

We would also like to mention the work [32] which considers a satisfaction game-based analysis for the MEC problem; however, it primarily focuses on meeting a minimum quality-of-service (QoS) requirement, rather than optimizing an objective function. In contrast, our work is centered on minimizing each user’s cost function in a fully non-cooperative setting, with the equilibrium solution formulated under the concept of Nash Equilibrium. Within the context of MEC systems, to the best of the authors’ knowledge, the only work employing the *mean-field type game* paradigm is [33] which is comprised of end-users offloading tasks, an intermediate task aggregator which pre-processes and stores all the tasks, and multiple ESs which can “pull” tasks for computation. Thus, decision making is carried out from the perspective of the ESs which decide on the amount of tasks that *they* can complete based on their energy and resource constraints. In contrast, in our current work, decision making is carried out at the end-user level, which decides on how much to execute locally and how much to offload, and without the presence of any intermediate authority like the task aggregator. Furthermore, our objective is based on the AoI based performance metric as opposed to [33] considering a penalty for the edge node based on how much edge resource is utilized. Thus, our work is the first to exploit the MFG and MM-MFG framework to serve time-critical MEC systems with and without priority access.

The main contributions of our work are as follows:

- 1) First, we provide a novel formulation of the computation offloading problem in an (*equitable access*) MEC system using AoI objective within the paradigm of non-cooperative game theory. This appropriately takes into account the selfish nature of the end-users and to better support time-critical applications in comparison to traditional performance metrics of delay and throughput (see for instance, [4], [5]). The novelty further lies in that our formulation allows for direct optimization of the AoI-based objective by being able to compute closed-form expressions for the average AoI (for a series-parallel server combination).
- 2) Second, to alleviate the associated issues of scalability and tractable Nash equilibrium computation within a large-user system (due to lack of policy information of other users), we employ the novel MFGs framework (with heterogeneous user types) to provide low complexity algorithms to compute *completely distributed approximate Nash equilibrium* offloading policies for the energy-constrained users. This plays a significant part in reducing a) the storage requirement, and b) communication overheads, due to information exchange between the large number of users.
- 3) We also extend the framework in 1) to a novel setting of MEC with priority access motivated by techniques in cognitive radio networks, where end-users are labeled as primary and secondary based on the priority of access

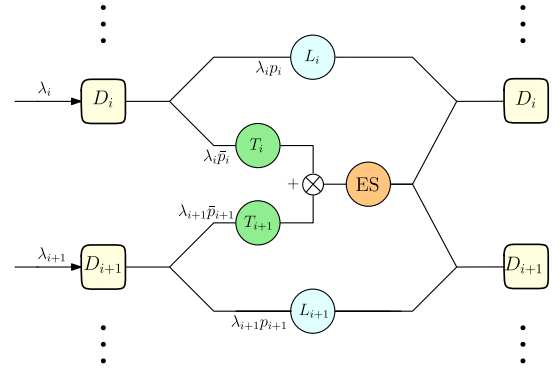


Fig. 2. The flow of incoming tasks is shown for a system of N devices. In particular, for device D_i : L_i and T_i denote the device’s local processor and its transmitter, resp.; ES denotes the edge server; p_i and $\bar{p}_i = 1 - p_i$ denote the Bernoulli probability according to which a task is either chosen to be served locally or offloaded to the ES. (Note that the ‘+’ symbol denotes the aggregation of task arrivals from different devices.).

given to them. For the former problem, we borrow the elegant framework of major-minor MFGs from the control theory literature to, again, provide low-complexity algorithms to compute distributed Nash equilibrium offloading policies for both the primary and the secondary users.

- 4) We finally present an extensive numerical analysis to validate our theoretical results and interestingly demonstrate that our formulation results in the well-studied $[O(V), O(1/V)]$ power-execution delay trade-off. Further, our MFG-based approach closely approximates Nash equilibrium policies with a significant advantage of being fully decentralized. Finally, we also demonstrate that with our primary-secondary setup, we can greatly improve the effective resource utilization by 37%.

Distinctively, our approach provides a very clean *recipe* on how to design offloading policies for MEC systems rather than heuristic-based ones. In the process, we also provide complementary results on the computation of average AoIs for systems with series-sum-parallel connections of servers. These extend the results of the works [34], [35] and would be of independent interest.

Notations: $[N] := \{1, \dots, N\}$ denotes the set of agents. We use the shorthand $\text{Poi}(a)$ and $\exp(a)$ to denote Poisson and exponential random variables, respectively. For a policy vector $a = [a_1, \dots, a_N]$, a_{-i} denotes the policy vector of all users other than i . Additionally, we have listed other key notations used in this work in Table I.

II. MEC WITH EQUITABLE ACCESS— PROBLEM SETUP

In this section, we start by formulating the MEC problem where each device has equitable (one-hop) access to the ES. A schematic of the system model is shown in Fig. 2 which comprises N users and one ES. Associated with each user is a type variable ϕ , which belongs to a finite set Φ , and is sampled according to a probability distribution $\mathbb{P}_N(\phi)$ with $\lim_{N \rightarrow \infty} \mathbb{P}_N(\phi) = \mathbb{P}(\phi), \forall \phi$. The notion of *type* of a user allows us to introduce heterogeneity within the MEC system where each

TABLE I

LIST OF KEY NOTATIONS (SECTION-WISE); THE SUBSCRIPT NOTATION z_ϕ REFERS TO A PROPERTY z OF USER WITH TYPE ϕ WHILE THE ANALOGOUS NOTATION z_i (WHICH IS ALSO USED BUT NOT LISTED ABOVE TO AVOID REPETITION) DENOTES A PROPERTY OF USER WITH INDEX i

Sections II, IV	
N	number of (secondary) devices
ES	edge server
ϕ	generic device of type ϕ
λ_ϕ	arrival rate for type ϕ device
p_ϕ	local processing probability for type ϕ device
$\mu_{1\phi}$	mean transmission power for type ϕ device
$\mu_{2\phi}$	local processing frequency for type ϕ device
$P_{\phi,max}$	transmission power bound for type ϕ device
$f_{\phi,max}$	local frequency bound for type ϕ device
Sections V, VII	
D_{S_ϕ}	secondary device
L_{S_ϕ}	ϕ^{th} secondary device's local processor
T_{S_ϕ}	ϕ^{th} secondary device's transmitter
P_{max}	transmission power bound for primary device
$f_{P,max}$	local frequency bound primary device
D_P	primary device
L_P	primary device's local processor
T_P	primary device's transmitter
λ_P	arrival rate for primary device
p_P	local processing probability for primary device
μ_{1P}	mean transmission power for primary device
μ_{2P}	local processing frequency for primary device
$\eta_P(\eta_\phi)$	device coefficient for primary (type ϕ) device

device belonging to a different type can have possibly different rates of arrival, service rates, and other device characteristics.

Next, each device D_i needs to execute tasks arriving to the device. To assist in the execution process, an ES is available. Thus, each device has two task processing options: (1) it can run its own local processor L_i at a certain frequency, or (2) it can use its transmitter T_i to offload the computation to the ES, as shown in Fig. 2. The inter-arrival times of tasks arriving at the i th device D_i are distributed as an $\exp(\lambda_i)$ random variable (r.v.) for all $i \in [N]$ [36], [37]. If D_i decides to carry out the tasks on L_i , then it can operate the processor at a frequency $\mu_{2i} \leq f_{i,max}$. The service time of L_i is distributed as an $\exp(\mu_{2i})$ r.v. Accordingly, the processing power used is $P_{\ell,i} = \eta_i \mu_{2i}^3$, where η_i is a positive constant denoting the processor's effective capacitance [13].

On the other hand, if D_i decides to offload the task to the ES, then it gets served sequentially by T_i to the ES and the ES uploads it back to the device after processing. The transmission rate of T_i is modeled as an $\exp(\mu_{1i})$ r.v. with μ_{1i} being the mean transmission power usage and $\mu_{1i} \leq P_{i,max}$. The service time of task processing at the ES is modeled as an $\exp(\mu_3)$ r.v. where $\mu_3 \gg \mu_{1i}, \mu_{2i}$. We employ the last-come-first-serve with preemption (LCFS-P) discipline² at all the servers (L_i, T_i , and

the ES) and assume that the downloading time of the processed task by the device is negligible.³

Since the effective service rates provided by L_i and the series path of T_i and the ES are heterogeneous, we employ the i.i.d. Bernoulli distributed random variable with a mean p_i to split the incoming Poisson process into two independent Poisson processes with respective means $\lambda_i p_i$ and $\lambda_i \bar{p}_i$, where $\bar{p}_i = 1 - p_i$ (as in Fig. 2). Such a Bernoulli splitting process has been widely employed in the literature in systems with heterogeneous parallel connection of servers [34]. Finally, we measure the freshness of processed information at the device using the average AoI metric which has been widely employed in the literature [5], [36], [39], [40], [41], [42], [43], [44]. Formally, the AoI at the receiver (which is the device itself in our case) is defined as the time elapsed at the receiving end since the latest delivered information packet was generated at the source.

Thus, the aim of each device D_i is to find optimal offloading policies (i.e., choosing the decision variables p_i, μ_{1i}, μ_{2i}) to serve a two-fold objective of: (1) minimizing the average AoI of the tasks, and (2) minimizing the power consumed during local processing and transmission. Since this is a multi-objective optimization problem, in the sequel we use the scalarization approach [45] to set up each device's problem. Let us define $\boldsymbol{\mu}_1 := [\mu_{11}, \dots, \mu_{1N}]$, $\boldsymbol{\mu}_2 := [\mu_{21}, \dots, \mu_{2N}]$ and $\mathbf{p} := [p_1, \dots, p_N]$. Then, the fraction of time that L_i is busy can be computed as $t_{L_i} = \lambda_i p_i / (\lambda_i p_i + \mu_{2i})$ and the fraction of time that T_i is busy can be computed as $t_{T_i} = \lambda_i \bar{p}_i / (\lambda_i \bar{p}_i + \mu_{1i})$. Consequently, each device $i \in [N]$ wishes to solve the following problem.

Problem 1 (N -user game problem): Each device $i \in [N]$ aims to minimize its cost $J^{N,i}$:

$$\begin{aligned} \min_{(p_i, \mu_{1i}, \mu_{2i}) \in [0,1] \times \mathbb{R}^2} \quad & J^{N,i}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \\ \text{s.t.} \quad & 0 \leq \mu_{1i} \leq P_{i,max} \\ & 0 \leq \mu_{2i} \leq f_{i,max}, \end{aligned}$$

$$J^{N,i}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2) := t_{T_i} \mu_{1i} + t_{L_i} \eta_i \mu_{2i}^3 + V_i \Delta_i^{(N)}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2).$$

Here, the first two terms in $J^{N,i}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ denote the average power consumed at T_i and L_i , respectively, and $\Delta_i^{(N)}(\cdot, \cdot, \cdot)$ denotes the average AoI incurred by D_i and $V_i > 0$ is the scalarization parameter which weights information freshness versus power consumption. A high value of V_i indicates that the device i cares about time responsiveness more than the power consumed and vice versa.

We note that the Problem 1 is a game problem due to the presence of other devices' policies in the cost optimization problem of the i th device. This requires each device to know

to compute average AoI expressions for a system with a hybrid connection of series-parallel servers.

³Motivated by autonomous vehicular systems or real-time monitoring systems, the uploaded tasks usually consist of high-quality images or videos which take non-negligible transmission duration versus the processed tasks, which constitute low-size commands (such as "accident ahead" signal, or the target's real-time position) which can be transmitted back to the IoT devices instantaneously. Also, the ES can be directly connected to the power source, and thus, it can use significantly higher transmission power. However, as IoT devices have finite batteries, their transmission times may not be negligible.

²The motivation behind using a preemption based discipline is two-fold: 1) it allows for efficient operation of systems with shared resources and selfish users (quantified using the price of anarchy or the price of stability metrics) as observed in the literature [38], and 2) it allows for a manageable state space

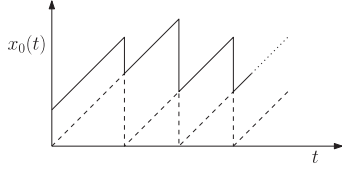


Fig. 3. Evolution of AoI at the receiver.

the policies of the other devices to compute its own, which can incur a significant communication overhead, especially in a large-user scenario. Thus, we will later employ the MFG framework to alleviate this issue and allow for tractable policy design. However, first, to complete the formulation of the above problem, we need to characterize the expression for the average AoI, $\Delta_i^{(N)}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, which we will derive in the next section. Also, henceforth, we refer to the triple $(p_i, \mu_{1i}, \mu_{2i})$ as the policy of device D_i .

III. AGE OF INFORMATION (AOI) CALCULATIONS

In this section, we will determine the average AoI for device i appearing in Problem 1. In this regard, we will use the method of stochastic hybrid systems (SHS) proposed in [46] first for continuous-state dynamical systems and then in [34], [47] for discrete-state modeling and AoI computations for different server constellations. For completeness, we first briefly review the main concepts in the next subsection. We refer the interested reader to [34] for details.

A. Description of SHS

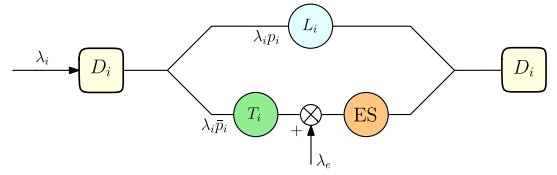
Let us briefly recall how an AoI process as shown in Fig. 3 can be modeled by a piecewise linear SHS. In any SHS, the overall state of the system can be described using a discrete-continuous pair $(s(t), x(t)) \in \mathcal{S} \times \mathbb{R}^{n+1}$ for all time $t \geq 0$, with $\mathcal{S}$ being a finite set. Here, $n+1$ denotes the number of servers involved in the constellation plus the monitor/receiver itself. The continuous state $x(t)$ evolves according to a stochastic differential equation

$$dx(t) = \mathbf{a}(t, s, x)dt + \mathbf{b}(t, s, x)dB(t), \quad (1)$$

where $B(t)$ is a standard Brownian motion, and $\mathbf{a}$ and $\mathbf{b}$ are real-valued mappings. Further, the discrete state $s(t)$ evolves from a state s to a state s' with transition intensity $q\delta_{s,s'}$ within the set $\mathcal{S}$. The notation $\delta_{s,s'}$ denotes the Kronecker delta function, which equals 1 if and only if $s = s'$, and 0 otherwise. For each such transition, the continuous state also jumps to a new value x' and is defined using a real-valued mapping $x' = \mathbf{h}(t, s, x; s')$. The resulting process $x(t)$ thus has piecewise continuous sample paths.

Let us restrict our attention now to systems in which $s(t)$ evolves as a finite-state Markov chain (FS-MC). Within such a setting, the AoI process can be characterized as a special case of the SHS theory which is a piecewise linear SHS with $\mathbf{a}(t, s, x) = \mathbf{u}_s \in \{0, 1\}$, $\mathbf{b}(t, s, x) = 0$, and $\mathbf{h}(t, s, x; s') = xA_s$, where $A_s \in \{0, 1\}^{(n+1) \times (n+1)}$.

Following [47], we define $\pi_{s'}(t) := \mathbb{P}(s(t) = s')$ as the discrete-Markov state probability of being at state $s(t) = s'$ and

Fig. 4. Task flow from the perspective of device D_i .

$v_{s'k}(t) := \mathbb{E}[x_k(t)\delta_{s(t),s'}]$, which measures the correlation between the age process $x_k(t)$ in server k with the discrete state $s(t)$ at timestep t . Further, let us denote the set of possible outgoing transitions from a particular state s as $\mathcal{L}_s := \{\ell : s_\ell = s\}$ and the set of possible incoming transitions to a state s' as $\mathcal{L}'_{s'} := \{\ell : s_\ell = s'\}$. Then, under the assumption of ergodicity of the FS-MC, a unique steady state distribution $\bar{\pi} := [\bar{\pi}_1, \dots, \bar{\pi}_m]$ exists [48] and satisfies the conservation law,

$$\bar{\pi}_s \sum_{\ell \in \mathcal{L}_s} q^\ell = \sum_{\ell' \in \mathcal{L}'_s} q^{\ell'} \bar{\pi}_{s_{\ell'}}, \quad \forall s \in \mathcal{S}, \quad (2a)$$

$$\sum_{s \in \mathcal{S}} \bar{\pi}_s = 1, \quad (2b)$$

where m denotes the cardinality of $\mathcal{S}$. Let us denote by v_s the vector $[v_{s0}, v_{s1}, \dots, v_{sn}]$ of correlations for all servers. We note that here and henceforth, we always take the index of the monitor (which in our case will be the device D_i) as 0, which means x_0 denotes the age process of the monitor and v_{s0} denotes the correlation function as defined above for the monitor. Consequently, one can obtain the following result.

Theorem 1 ([47, Theorem 4]): Suppose that $\bar{\pi}$ is the state distribution of the FS-MC and there exists a stationary solution $\bar{v} := [\bar{v}_1, \dots, \bar{v}_m]$ of the process $v(t)$ satisfying

$$\bar{v}_s \sum_{\ell \in \mathcal{L}_s} q^\ell = \mathbf{u}_s \bar{\pi}_s + \sum_{\ell' \in \mathcal{L}'_s} q^{\ell'} \bar{v}_{s_{\ell'}} A_{\ell'}. \quad (3)$$

Then, the average AoI is given by $\Delta := \sum_{s \in \mathcal{S}} \bar{v}_{s0}$.

We will use the above theorem to compute $\Delta_i^{(N)}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ as defined in Problem 1.

B. Average AoI Calculation

In this subsection, we will provide an approximate average AoI expression for device D_i by assuming that the incoming Poisson process has a sufficiently large parameter $\lambda_i, \forall i$. The justification for carrying out an approximate analysis is due to exponential increase in the state space of the FS-MC (defined in the previous subsection), the details of which will be provided later in Remark 2. Let us begin by concentrating on the incoming task flow from the perspective of device D_i as shown in Fig. 4. Under the assumption that the departure process of T_j 's for all users closely follows a Poisson process, the exogenous incoming rate (referred to as λ_e in Fig. 4) interfering with the AoI process of D_i approximately follows a $\text{Poi}(\lambda_e)$ distribution [47], where $\lambda_e := \sum_{j=1, j \neq i}^N \frac{\lambda_j \bar{p}_i \mu_{1,i}}{\lambda_j \bar{p}_i + \mu_{1,i}}$. The quantity λ_e denotes the cumulative exogenous rate of users other than the i th one.

Next, let us mathematically formulate the AoI process for device D_i using the SHS method as discussed in the previous subsection. To this end, we need to define the state space $\mathcal{S}$

TABLE II
STATE DICTIONARY FOR THE FINITE FS-MC

state	server 1 (T_i)	server 2 (L_i)	server 3 (ES)
s_1	freshest	2^{nd} freshest	oldest
s_2	freshest	oldest	2^{nd} freshest
s_3	2^{nd} freshest	freshest	oldest
s_4	<i>no packet</i>	freshest	2^{nd} freshest
s_5	<i>no packet</i>	2^{nd} freshest	freshest
s_6	<i>no packet</i>	freshest	class 2
s_7	freshest	2^{nd} freshest	class 2
s_8	2^{nd} freshest	freshest	class 2

along with its corresponding transition functions to characterize the FS-MC, which is done as follows. Also, henceforth, we will refer to the exogeneous packets as in Fig. 4 as that of class 2.

1) *State Space*: The state space S of the FS-MC constitutes 8 states which keep track of the server that is currently servicing the freshest, the second freshest, and the oldest packets of D_i and the packets of class 2. Detailed descriptions are provided in Table II. For example, state s_4 denotes that for device D_i , server T_i is idling, and servers L_i and the ES are servicing the freshest and the 2nd freshest packet of D_i , respectively. The choice of the states is made in view of the concept of fake updates [34], whereby without any loss of generality, we can assume that all servers which do not precede a node of packet arrival (L_i and ES in our case) are busy all the time by running a fake packet in that server. The type and current AoI of the fake packet is the same as the departing packet. Consequently, we have that $u_s = [1 \ 1 \ 1 \ 1]$ for $s = s_1, s_2, s_3, s_7, s_8$ and $u_s := \hat{u}_s = [1 \ 0 \ 1 \ 1]$ for $s = s_4, s_5, s_6$ since the transmitter T_i is idling in the latter three states.

Remark 1: It is essential to take care of the emphasized phrase in the later part of the previous paragraph on which server can run a fake packet. This is because the transmitter T_i in our formulation precedes the point of arrival of exogeneous packets (as in Fig. 4). Thus, the SHS model should take into account whether it is idling or is busy. This prevents us from running a fake update at T_i and we explicitly account for the idle state of T_i by writing “no packet” in Table II.

2) *Transition Functions*: Now that we have completely described the state S , we next list the possible transitions in the FS-MC in Table III. Alongside, we also track the current AoIs of the packets in each server after each transition, which is listed as the AoI vector $x'(t) := [x'_0(t) \ x'_1(t) \ x'_2(t) \ x'_3(t)]$, where $x'_0(t)$, $x'_1(t)$, $x'_2(t)$, and $x'_3(t)$ denote the AoI at D_i , the local processor L_i , the transmitter T_i and the ES, respectively. Note that henceforth, we omit the subscript index i for brevity.

Now that the FS-MC is completely characterized by Tables II and III, we now proceed towards computing the average AoI for D_i . Let us begin by defining the quantities $a := \lambda + \lambda_e + \mu_1 + \mu_2 + \mu_3$ and $\hat{a} := a - \mu_1$. Then, using (2), the steady state probability vector $\bar{\pi}$ satisfies (2b) and the following set of equations in (4a)–(4h):

$$a\bar{\pi}_1 = (\lambda\bar{p} + \mu_2 + \mu_3)\bar{\pi}_1 + \lambda\bar{p}(\bar{\pi}_3 + \bar{\pi}_4) \quad (4a)$$

$$a\bar{\pi}_2 = (\lambda\bar{p} + \mu_2 + \mu_3)\bar{\pi}_2 + \lambda\bar{p}\bar{\pi}_5 \quad (4b)$$

TABLE III
STATE TRANSITIONS OF THE FS-MC AND ASSOCIATED AOI JUMPS

s	q	s'	$x' = xA_s$	$v_s A_s$
s_1	λp	s_3	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{10} \ \bar{v}_{11} \ 0 \ \bar{v}_{13}]$
	$\lambda\bar{p}$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{10} \ 0 \ \bar{v}_{12} \ \bar{v}_{13}]$
	λ_e	s_7	$[x_0 \ x_1 \ x_2 \ x_0]$	$[\bar{v}_{10} \ \bar{v}_{11} \ \bar{v}_{12} \ \bar{v}_{10}]$
	μ_1	s_5	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{10} \ 0 \ \bar{v}_{12} \ \bar{v}_{11}]$
	μ_2	s_1	$[x_2 \ x_1 \ x_2 \ x_2]$	$[\bar{v}_{12} \ \bar{v}_{11} \ \bar{v}_{12} \ \bar{v}_{12}]$
	μ_3	s_1	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{13} \ \bar{v}_{11} \ \bar{v}_{12} \ \bar{v}_{13}]$
s_2	λp	s_3	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{20} \ \bar{v}_{21} \ 0 \ \bar{v}_{23}]$
	$\lambda\bar{p}$	s_2	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{20} \ 0 \ \bar{v}_{22} \ \bar{v}_{23}]$
	λ_e	s_7	$[x_0 \ x_1 \ x_2 \ x_0]$	$[\bar{v}_{20} \ \bar{v}_{21} \ \bar{v}_{22} \ \bar{v}_{20}]$
	μ_1	s_5	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{20} \ 0 \ \bar{v}_{22} \ \bar{v}_{21}]$
	μ_2	s_2	$[x_2 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{22} \ \bar{v}_{21} \ \bar{v}_{22} \ \bar{v}_{23}]$
	μ_3	s_2	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{23} \ \bar{v}_{21} \ \bar{v}_{22} \ \bar{v}_{23}]$
s_3	λp	s_3	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{30} \ \bar{v}_{31} \ 0 \ \bar{v}_{33}]$
	$\lambda\bar{p}$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{30} \ 0 \ \bar{v}_{32} \ \bar{v}_{33}]$
	λ_e	s_8	$[x_0 \ x_1 \ x_2 \ x_0]$	$[\bar{v}_{30} \ \bar{v}_{31} \ \bar{v}_{32} \ \bar{v}_{30}]$
	μ_1	s_4	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{30} \ 0 \ \bar{v}_{32} \ \bar{v}_{31}]$
	μ_2	s_3	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{32} \ \bar{v}_{32} \ \bar{v}_{32} \ \bar{v}_{32}]$
	μ_3	s_3	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{33} \ \bar{v}_{31} \ \bar{v}_{32} \ \bar{v}_{33}]$
s_4	λp	s_4	$[x_0 \ 0 \ 0 \ x_3]$	$[\bar{v}_{40} \ 0 \ 0 \ \bar{v}_{43}]$
	$\lambda\bar{p}$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{40} \ 0 \ \bar{v}_{42} \ \bar{v}_{43}]$
	λ_e	s_6	$[x_0 \ 0 \ x_2 \ x_0]$	$[\bar{v}_{40} \ 0 \ \bar{v}_{42} \ \bar{v}_{40}]$
	μ_2	s_4	$[x_2 \ 0 \ x_2 \ x_2]$	$[\bar{v}_{42} \ 0 \ \bar{v}_{42} \ \bar{v}_{42}]$
	μ_3	s_4	$[x_3 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{43} \ 0 \ \bar{v}_{42} \ \bar{v}_{43}]$
s_5	λp	s_4	$[x_0 \ 0 \ 0 \ x_3]$	$[\bar{v}_{50} \ 0 \ 0 \ \bar{v}_{53}]$
	$\lambda\bar{p}$	s_2	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{50} \ 0 \ \bar{v}_{52} \ \bar{v}_{53}]$
	λ_e	s_6	$[x_0 \ 0 \ x_2 \ x_0]$	$[\bar{v}_{50} \ 0 \ \bar{v}_{52} \ \bar{v}_{50}]$
	μ_2	s_5	$[x_2 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{52} \ 0 \ \bar{v}_{52} \ \bar{v}_{53}]$
	μ_3	s_5	$[x_3 \ 0 \ x_3 \ x_3]$	$[\bar{v}_{53} \ 0 \ \bar{v}_{53} \ \bar{v}_{53}]$
s_6	λp	s_6	$[x_0 \ 0 \ 0 \ x_3]$	$[\bar{v}_{60} \ 0 \ 0 \ \bar{v}_{63}]$
	$\lambda\bar{p}$	s_7	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{60} \ 0 \ \bar{v}_{62} \ \bar{v}_{63}]$
	λ_e	s_6	$[x_0 \ 0 \ x_2 \ x_0]$	$[\bar{v}_{60} \ 0 \ \bar{v}_{62} \ \bar{v}_{60}]$
	μ_2	s_6	$[x_2 \ 0 \ x_2 \ x_2]$	$[\bar{v}_{62} \ 0 \ \bar{v}_{62} \ \bar{v}_{62}]$
	μ_3	s_6	$[x_3 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{63} \ 0 \ \bar{v}_{62} \ \bar{v}_{63}]$
s_7	λp	s_8	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{70} \ \bar{v}_{71} \ 0 \ \bar{v}_{73}]$
	$\lambda\bar{p}$	s_7	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{70} \ 0 \ \bar{v}_{72} \ \bar{v}_{73}]$
	λ_e	s_7	$[x_0 \ x_1 \ x_2 \ x_0]$	$[\bar{v}_{70} \ \bar{v}_{71} \ \bar{v}_{72} \ \bar{v}_{70}]$
	μ_1	s_5	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{70} \ 0 \ \bar{v}_{72} \ \bar{v}_{71}]$
	μ_2	s_7	$[x_2 \ x_1 \ x_2 \ x_2]$	$[\bar{v}_{72} \ \bar{v}_{71} \ \bar{v}_{72} \ \bar{v}_{72}]$
	μ_3	s_7	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{73} \ \bar{v}_{71} \ \bar{v}_{72} \ \bar{v}_{73}]$
s_8	λp	s_8	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{80} \ \bar{v}_{81} \ 0 \ \bar{v}_{83}]$
	$\lambda\bar{p}$	s_7	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{80} \ 0 \ \bar{v}_{82} \ \bar{v}_{83}]$
	λ_e	s_8	$[x_0 \ x_1 \ x_2 \ x_0]$	$[\bar{v}_{80} \ \bar{v}_{81} \ \bar{v}_{82} \ \bar{v}_{80}]$
	μ_1	s_4	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{80} \ 0 \ \bar{v}_{82} \ \bar{v}_{81}]$
	μ_2	s_8	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{82} \ \bar{v}_{82} \ \bar{v}_{82} \ \bar{v}_{82}]$
	μ_3	s_8	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{83} \ \bar{v}_{81} \ \bar{v}_{82} \ \bar{v}_{83}]$

$$a\bar{\pi}_3 = (\lambda p + \mu_2 + \mu_3)\bar{\pi}_3 + \lambda p(\bar{\pi}_1 + \bar{\pi}_2) \quad (4c)$$

$$\hat{a}\bar{\pi}_4 = (\lambda p + \mu_2 + \mu_3)\bar{\pi}_4 + \lambda p\bar{\pi}_5 + \mu_1(\bar{\pi}_3 + \bar{\pi}_8) \quad (4d)$$

$$\hat{a}\bar{\pi}_5 = (\mu_2 + \mu_3)\bar{\pi}_5 + \mu_1(\bar{\pi}_1 + \bar{\pi}_2 + \bar{\pi}_7) \quad (4e)$$

$$\hat{a}\bar{\pi}_6 = (\lambda p + \lambda_e + \mu_2 + \mu_3)\bar{\pi}_6 + \lambda_e(\bar{\pi}_4 + \bar{\pi}_5) \quad (4f)$$

$$a\bar{\pi}_7 = (\lambda\bar{p} + \lambda_e + \mu_2 + \mu_3)\bar{\pi}_7 + \lambda_e(\bar{\pi}_1 + \bar{\pi}_2) + \lambda\bar{p}(\bar{\pi}_6 + \bar{\pi}_8) \quad (4g)$$

$$a\bar{\pi}_8 = (\lambda p + \lambda_e + \mu_2 + \mu_3)\bar{\pi}_7 + \lambda_e\bar{\pi}_3 + \lambda p\bar{\pi}_7, \quad (4h)$$

which can further be equivalently written using a compact matrix form as $A_m \bar{\pi} = M_m \bar{\pi}$, where the entries A_m and M_m can be easily determined using (4) above.

The above set of linear equations (4) allows us to compute the distribution $\bar{\pi}$ by combining with (2b). Then, to compute the average AoI $\Delta_i^{(N)}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, it remains to compute the steady state vector $\bar{v}$ in Theorem 1 and then apply the formula for Δ in the theorem. To compute $\bar{v}$, we use (3) in Theorem 1 to write down the set of linear equations satisfied by its components in (5) (which again, can be equivalently written in compact form as $A_m \bar{v} = V_m \bar{v}$ with the entries of V_m determined using (5)). Consequently, the average AoI, $\Delta_i^{(N)}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, can be computed by first computing $\bar{\pi}$ using (4a)–(4h), substituting the same in (5a)–(5h) shown at the bottom of this page, and solving the latter set of equations. We summarize the result in the following theorem by resuming the use of subscript i corresponding to device D_i .

Theorem 2: Suppose that the arrival rate at device D_i is distributed as $\text{Poi}(\lambda_i)$ and its service rates as $\exp(\mu_{1i})$ and $\exp(\mu_{2i})$. Let the service rate of the ES be distributed as $\exp(\mu_3)$. Then, the average AoI $\Delta_i^{(N)}(\mathbf{p}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ exists and is given by solving (4) and (5).

We next note here that even though one can obtain the closed-form expressions for the average AoI using Theorem 2, long expressions preclude us from writing them in the paper. Furthermore, later, we will provide an algorithm to compute the equilibrium policies for all the devices, where we would not require symbolic expressions of the AoI but only a “function call” to solve a linear program, which can then be conveniently solved using any linear program solver.

We have thus completed the average AoI calculations, and now state the following important remark on how the approximate calculations performed in this subsection lead to tractable computations of the average AoI in the presence of a large number of users.

Remark 2: To understand the importance of the large λ assumption, let us focus, for simplicity, on a two-user case (i.e., when $N = 2$) and the perspective of device D_1 as shown in Fig. 5. Using this, one can perform the exact AoI computations

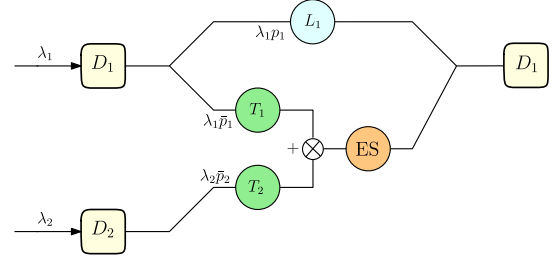


Fig. 5. Task flow for a two-user MEC system with one ES from the perspective of device D_1 .

since the incoming arrivals of both packets at devices D_1 and D_2 are Poisson, and hence, we can apply the SHS theory without any assumption on λ_i . In Fig. 4, however, this is not the case since the exogenous process preceding the ES is not Poisson, which makes the underlying finite state evolution of $s(t)$ non-Markovian and hence, precludes the application of the SHS theory. However, under the large λ approximation, the exogenous process closely follows a Poisson process, in which case, we can readily apply the SHS theory. Now, one could ask the question as to why not proceed with the set-up of Fig. 5. The reason for that is the exponential explosion of the state space of the underlying FS-MC with the number of users N . It is easy to see that one requires 2^{N-1} times the number of states currently required (which is 8) to characterize the FS-MC completely because one would need to track whether the transmitter T_j of the other devices is busy or idle. Thus, approximate analysis (as we carried out with large λ_i 's and summarized in Theorem 2) allows us to compute tractable AoI expressions with the state space being independent of the number of users.

With the above computations, we have now completely characterized Problem 1. In the next section, we proceed toward computing completely distributed Nash equilibrium (NE) solutions to Problem 1 using the MFG approach.

$$a\bar{v}_1 = \mathbf{u}_s \bar{\pi}_1 + \lambda \bar{p}[\bar{v}_{10} 0 \bar{v}_{12} \bar{v}_{13}] + \mu_2[\bar{v}_{12} \bar{v}_{11} \bar{v}_{12} \bar{v}_{12}] + \mu_3[\bar{v}_{13} \bar{v}_{11} \bar{v}_{12} \bar{v}_{13}] + \lambda \bar{p}[\bar{v}_{30} 0 \bar{v}_{32} \bar{v}_{33}] + \lambda \bar{p}[\bar{v}_{40} 0 \bar{v}_{42} \bar{v}_{43}], \quad (5a)$$

$$a\bar{v}_2 = \mathbf{u}_s \bar{\pi}_2 + \lambda \bar{p}[\bar{v}_{20} 0 \bar{v}_{22} \bar{v}_{23}] + \mu_2[\bar{v}_{22} \bar{v}_{21} \bar{v}_{22} \bar{v}_{23}] + \mu_3[\bar{v}_{23} \bar{v}_{21} \bar{v}_{23} \bar{v}_{23}] + \lambda \bar{p}[\bar{v}_{50} 0 \bar{v}_{52} \bar{v}_{53}], \quad (5b)$$

$$a\bar{v}_3 = \mathbf{u}_s \bar{\pi}_3 + \lambda \bar{p}[\bar{v}_{30} \bar{v}_{31} 0 \bar{v}_{33}] + \mu_2[\bar{v}_{32} \bar{v}_{32} \bar{v}_{32} \bar{v}_{32}] + \mu_3[\bar{v}_{33} \bar{v}_{31} \bar{v}_{32} \bar{v}_{33}] + \lambda \bar{p}[\bar{v}_{10} \bar{v}_{11} 0 \bar{v}_{13}] + \lambda \bar{p}[\bar{v}_{20} \bar{v}_{21} 0 \bar{v}_{23}], \quad (5c)$$

$$\hat{a}\bar{v}_4 = \hat{\mathbf{u}}_s \bar{\pi}_4 + \lambda \bar{p}([\bar{v}_{40} 0 0 \bar{v}_{43}] + [\bar{v}_{50} 0 0 \bar{v}_{53}]) + \mu_2[\bar{v}_{42} 0 \bar{v}_{42} \bar{v}_{42}] + \mu_3[\bar{v}_{43} 0 \bar{v}_{42} \bar{v}_{43}] + \mu_1([\bar{v}_{30} 0 \bar{v}_{32} \bar{v}_{31}] + [\bar{v}_{80} 0 \bar{v}_{82} \bar{v}_{81}]), \quad (5d)$$

$$\hat{a}\bar{v}_5 = \hat{\mathbf{u}}_s \bar{\pi}_5 + \mu_1[\bar{v}_{70} 0 \bar{v}_{72} \bar{v}_{71}] + \mu_2[\bar{v}_{52} 0 \bar{v}_{52} \bar{v}_{53}] + \mu_3[\bar{v}_{53} 0 \bar{v}_{53} \bar{v}_{53}] + \mu_1[\bar{v}_{10} 0 \bar{v}_{12} \bar{v}_{11}] + \mu_1[\bar{v}_{20} 0 \bar{v}_{22} \bar{v}_{21}], \quad (5e)$$

$$\hat{a}\bar{v}_6 = \hat{\mathbf{u}}_s \bar{\pi}_6 + \lambda \bar{p}[\bar{v}_{60} 0 0 \bar{v}_{63}] + \mu_2[\bar{v}_{62} 0 \bar{v}_{62} \bar{v}_{63}] + \mu_3[\bar{v}_{63} 0 \bar{v}_{62} \bar{v}_{63}] + \lambda_e([\bar{v}_{40} 0 \bar{v}_{42} \bar{v}_{40}] + [\bar{v}_{50} 0 \bar{v}_{52} \bar{v}_{50}] + [\bar{v}_{60} 0 \bar{v}_{62} \bar{v}_{60}]), \quad (5f)$$

$$a\bar{v}_7 = \mathbf{u}_s \bar{\pi}_7 + \lambda \bar{p}[\bar{v}_{60} 0 \bar{v}_{62} \bar{v}_{63}] + \lambda \bar{p}[\bar{v}_{70} 0 \bar{v}_{72} \bar{v}_{73}] + \lambda \bar{p}[\bar{v}_{80} 0 \bar{v}_{82} \bar{v}_{83}] + \lambda_e[\bar{v}_{10} \bar{v}_{11} \bar{v}_{12} \bar{v}_{10}] + \lambda_e[\bar{v}_{20} \bar{v}_{21} \bar{v}_{22} \bar{v}_{20}] + \lambda_e[\bar{v}_{70} \bar{v}_{71} \bar{v}_{72} \bar{v}_{70}] + \mu_3[\bar{v}_{73} \bar{v}_{71} \bar{v}_{72} \bar{v}_{73}] + \mu_2[\bar{v}_{72} \bar{v}_{71} \bar{v}_{72} \bar{v}_{72}], \quad (5g)$$

$$a\bar{v}_8 = \mathbf{u}_s \bar{\pi}_8 + \lambda \bar{p}([\bar{v}_{70} \bar{v}_{71} 0 \bar{v}_{73}] + [\bar{v}_{80} \bar{v}_{81} 0 \bar{v}_{83}]) + \lambda_e([\bar{v}_{30} \bar{v}_{31} \bar{v}_{32} \bar{v}_{30}] + [\bar{v}_{80} \bar{v}_{81} \bar{v}_{82} \bar{v}_{80}]) + \mu_2[\bar{v}_{82} \bar{v}_{82} \bar{v}_{82} \bar{v}_{82}] + \mu_3[\bar{v}_{83} \bar{v}_{81} \bar{v}_{82} \bar{v}_{83}]. \quad (5h)$$

IV. MEAN-FIELD GAME (MFG)

Now that we have completely characterized the non-cooperative game problem defined in Problem 1, our aim is to compute NE policies $(p_i^*, \mu_{1i}^*, \mu_{2i}^*)$, $\forall i$, which satisfy the following set of inequalities [49],

$$\begin{aligned} & J^{N,i}(p_i^*, \mu_{1i}^*, \mu_{2i}^*, p_{-i}^*, \mu_{1,-i}^*, \mu_{2,-i}^*) \\ & \leq J^{N,i}(p_i, \mu_{1i}, \mu_{2i}, p_{-i}^*, \mu_{1,-i}^*, \mu_{2,-i}^*), \forall i \in [N], \end{aligned} \quad (6)$$

where the notation x_{-i} stands for the vector of variables x_j for all users j excluding user i . Briefly, the above set of inequalities state that any *rational* user who tries to deviate from the NE policy (p^*, μ_1^*, μ_2^*) incurs a higher cost, and thus it is in the best interest of each user to follow the NE policy. However, since the cost function of each user depends on the policies of the other users, no user can independently optimize to compute its own policy. Hence, each user requires the knowledge of the policies/parameters of the others in the population, which (a) can be difficult to acquire particularly within a large user setting, and (b) can incur significant communication overhead. Thus, to alleviate this issue, we would like to design completely distributed NE policies for the users where *each of them utilizes only their own local information* and the statistical information of the system (such as the limiting distribution $\mathbb{P}(\phi)_{\forall \phi \in \Phi}$). In this regard, we leverage the elegant framework of MFGs [8], [27], [31].

Motivated by statistical physics, the theory of MFGs aims to approximate the complex interactions within an interacting particle system with an *average effective field* (also referred to as the mean-field) generated by the corresponding infinite particle system. under suitable assumptions of “symmetry”, “indistinguishability”, and “anonymity”. This creates a *decoupling* effect and each particle now best reacts to the *mean-field* generated by the entire population in a manner such that its own behavior is consistent with that of the other particles without worrying about the policies of other particles within the population, where “consistency” is ensured through the solution of a fixed-point equation. Consequently, to derive equilibrium solutions, it suffices to consider the perspective of one generic particle, which represents the entire population. With the above prelude on MFGs, let us now set up the generic device’s optimization problem in the next subsection.

A. Generic Device Optimization Problem

Let us begin by focusing on a generic device of type $\phi \in \Phi$. The tasks arriving at the generic device D_ϕ are distributed as a $\text{Poi}(\lambda_\phi)$ r.v. which are sent to either its local processor or to the transmitter for offloading by employing a mean p_ϕ i.i.d. Bernoulli distributed r.v. This divides the incoming arrival process into two independent Poisson processes with respective means $\lambda_\phi p_\phi$ and $\lambda_\phi \bar{p}_\phi$. In addition, the service times of the generic transmitter and the generic local processor are distributed as exponential r.v.s with parameters $\mu_{1,\phi}$ and $\mu_{2,\phi}$, respectively. The corresponding upper bounds on the service rates are denoted as $P_{\phi,max}$ and $f_{\phi,max}$, respectively.

Next, let us introduce the following quantities $\rho^{(N)} := \frac{\lambda_e}{N\mu_3}$ and $\rho := \lim_{N \rightarrow \infty} \rho^{(N)}$, where $\rho^{(N)}$ denotes the mean load on the ES and ρ denotes the infinite user (or the mean-field (MF)) approximation as discussed earlier. Then, we have that for a large user MEC system, the exogenous arrival rate λ_e can be approximated as

$$\lambda_e = (N-1)\mu_3 \times \lambda_e / (N-1)\mu_3 \approx (N-1)\mu_3 \rho. \quad (7)$$

Consequently, the average AoI of the generic device becomes

$$\Delta_\phi(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}, \rho) := \Delta_\phi^{(N)}(p, \mu_1, \mu_2, \lambda_e) |_{\lambda_e = (N-1)\rho\mu_3}, \quad (8)$$

where the notation $x(z) |_{z=a}$ denotes the value of x when a is substituted for the argument z . Then, we can formally state the generic device’s optimization problem as follows.

Problem 2 (Generic device optimization problem):

$$\begin{aligned} & \min_{(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}) \in [0,1] \times \mathbb{R}^2} J_\rho(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}) \\ & \text{s.t. } 0 \leq \mu_{1,\phi} \leq P_{\phi,max} \\ & \quad 0 \leq \mu_{2,\phi} \leq f_{\phi,max}, \end{aligned}$$

where $J_\rho(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}) := t_{T_\phi} \mu_{1,\phi} + t_{L_\phi} \eta_\phi \mu_{2,\phi}^3 + V_\phi \Delta_\phi(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}, \rho)$, with $t_{L_\phi} = \lambda_\phi p_\phi / (\lambda_\phi p_\phi + \mu_{2,\phi})$, $t_{T_\phi} = \lambda_\phi \bar{p}_\phi / (\lambda_\phi \bar{p}_\phi + \mu_{1,\phi})$ and $\Delta_\phi(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}, \rho)$ is defined in (8).

Consequently, the MFG is defined using two operators, namely the optimality and the consistency operators as:

- 1) Optimality for generic user of type ϕ , $\forall \phi \in \Phi$:

$$(\hat{p}_\phi, \hat{\mu}_{1,\phi}, \hat{\mu}_{2,\phi}) = \Psi_1(\rho) := \text{argmin}_{(p_\phi, \mu_{1,\phi}, \mu_{2,\phi})} J_\rho(p_\phi, \mu_{1,\phi}, \mu_{2,\phi})$$

subject to the constraints in Problem 2.

- 2) Consistency of the mean-field:

$$\hat{\rho} = \Psi_2(\hat{p}_\phi, \hat{\mu}_{1,\phi}, \hat{\mu}_{2,\phi}) := \frac{1}{\mu_3} \mathbb{E}_{\mathbb{P}(\phi)} \left[\frac{\lambda_\phi \hat{p}_\phi \hat{\mu}_{1,\phi}}{\lambda_\phi \hat{p}_\phi + \hat{\mu}_{1,\phi}} \right].$$

Briefly, the optimality operator $\Psi_1(\cdot)$ outputs an optimal policy $(p_\phi, \mu_{1,\phi}, \mu_{2,\phi})$ for a user of type ϕ for a given value of ρ . Furthermore, the consistency operator $\Psi_2(\cdot)$ generates a new ρ by using the above optimal policy $(p_\phi, \mu_{1,\phi}, \mu_{2,\phi})$ and captures the fact that the generic user should act in a way such that its behavior is *consistent* with the load generated at the ES. The mean-field equilibrium (MFE) which constitutes the tuple of equilibrium policies for all types $(p_{\phi,MFE}, \mu_{1,\phi,MFE}, \mu_{2,\phi,MFE})_{\forall \phi}$ and the equilibrium mean-field (ρ_{MFE}) , is given by the fixed point of the composite map of Ψ_1 and Ψ_2 . To compute the MFE of the MEC game, we next provide a low complexity Algorithm 1, which is based on the technique of projected block-coordinate gradient-descent.

In summary, Algorithm 1 computes a fixed point of the composite operator $\Psi_2 \circ \Psi_1$. Thus, we start by randomly initialization the policy and the mean-field (as in line 4 of the algorithm). Then, given a value of ρ , we solve all generic users’ optimization problems defined in Problem 2 using block coordinate gradient descent method [50] (lines 8-15). Subsequently, we update the mean-field via Krasnoselskij’s iteration using the optimal policy obtained (line 27). The MFE is then given by the final iterate (line 30). We also note here that the main advantage of the above proposed MFG approach is that once we compute the

Algorithm 1: Fixed-Point Iteration for Computing MFE Policy of a Generic Device.

```

1: Input:  $V_\phi, \eta_\phi, \lambda_\phi, \mu_3, \forall \phi$  # System parameters
2: Input:  $\varepsilon_1, \varepsilon_2$  # tolerance parameters
3: Input:  $\gamma_1, \gamma_2$  # Iteration step sizes
4: Initialize:  $\hat{\rho}^{(0)}, \sigma_\phi^{(0)} := (p_\phi^{(0)}, \mu_{1,\phi}^{(0)}, \mu_{2,\phi}^{(0)}), \forall \phi$ 
5:  $k \leftarrow 1$ 
6: while  $|\hat{\rho}^{(m)} - \hat{\rho}^{(m-1)}| > \varepsilon_1$  do
7:   for  $\phi \in \Phi$  do
8:      $\ell \leftarrow 1$ 
9:     while  $|\hat{\sigma}_\phi^{(m')}(1) - \hat{\sigma}_\phi^{(m'-1)}(1)| > \varepsilon_2$ 
10:       $\hat{\sigma}_\phi^{(\ell)}(1) \leftarrow$ 
11:       $\hat{\sigma}_\phi^{(\ell-1)}(1) - \gamma_2 \nabla J_{\hat{\rho}^{(k-1)}}(p_\phi^{(\ell-1)}, \mu_{1,\phi}^{(k-1)}, \mu_{2,\phi}^{(k-1)})$ 
12:       $\ell \leftarrow \ell + 1$ 
13:    end while
14:     $p_\phi^{(k)} = \hat{\sigma}_\phi^{(m')}(1)$ 
15:     $\ell \leftarrow 1$ 
16:    while  $|\hat{\sigma}_\phi^{(m')}(2) - \hat{\sigma}_\phi^{(m'-1)}(2)| > \varepsilon_2$  do
17:       $\hat{\sigma}_\phi^{(\ell)}(2) \leftarrow$ 
18:       $\hat{\sigma}_\phi^{(\ell-1)}(2) - \gamma_2 \nabla J_{\hat{\rho}^{(k-1)}}(p_\phi^{(k)}, \mu_{1,\phi}^{(\ell-1)}, \mu_{2,\phi}^{(k-1)})$ 
19:       $\ell \leftarrow \ell + 1$ 
20:    end while
21:     $\mu_{1,\phi}^{(k)} = \hat{\sigma}_\phi^{(m')}(2)$ 
22:     $\ell \leftarrow 1$ 
23:    While  $|\hat{\sigma}_\phi^{(m')}(3) - \hat{\sigma}_\phi^{(m'-1)}(3)| > \varepsilon_2$  do
24:       $\hat{\sigma}_\phi^{(\ell)}(3) \leftarrow$ 
25:       $\hat{\sigma}_\phi^{(\ell-1)}(3) - \gamma_2 \nabla J_{\hat{\rho}^{(k-1)}}(p_\phi^{(k)}, \mu_{1,\phi}^{(k)}, \mu_{2,\phi}^{(\ell-1)})$ 
26:       $\ell \leftarrow \ell + 1$ 
27:    end while
28:     $\mu_{2,\phi}^{(k)} = \hat{\sigma}_\phi^{(m')}(3)$ 
29:  end for
30:   $\hat{\rho}^{(k)} \leftarrow (1 - \gamma_1)\hat{\rho}^{(k-1)} + \gamma_1 \mathbb{E}_{\mathbb{P}(\phi)} \left[ \frac{\lambda_\phi(1 - \hat{p}_\phi^{(k)})\hat{\mu}_{1,\phi}^{(k)}}{\mu_3(\lambda_\phi(1 - \hat{p}_\phi^{(k)}) + \hat{\mu}_{1,\phi}^{(k)})} \right]$ 
31:   $k \leftarrow k + 1$ 
32: end while
33: Output:  $\hat{\rho}^{(m)}, \sigma_\phi^{(m)}, \forall \phi$ .

```

MF ρ offline, when the users operate in real-time, they would be able to make decisions based on only local information and the pre-computed MF. This shows the fully distributed nature of the MF approach as opposed to the NE based approaches in the literature to design offloading policies.

We also note that a simple computation of the Hessian of J_ρ reveals its highly non-convex nature in the policy of the device of type ϕ , which makes the resulting MFG, a non-convex game. This means that there could possibly exist multiple MFGs for the aforementioned game. Thus, in search of the best one, we run Algorithm 1 for multiple random initializations, and subsequently pick the best (in the case where the resultant MFEs are comparable). Finally, to see how the MFG approach which is based on the infinite-user approximation of the finite-user system performs, we will later compare it with the NE of the MEC game computed using a best response dynamics algorithm (see

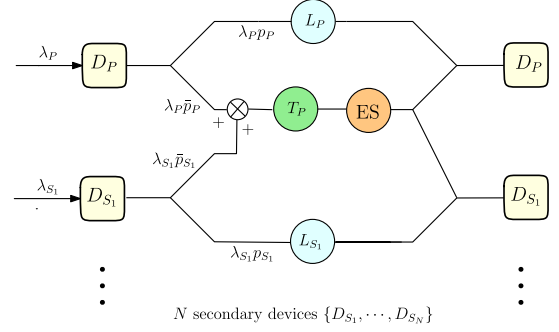


Fig. 6. Task flow for the primary device D_P and the i th secondary device D_{S_i} .

Algorithm 3 in the Appendix). We note here that Algorithm 1 requires running gradient computations for only the number of types of users rather than that of all the users, which is typically much less than the number of users in the MEC system. This is because users of the same type adopt the same policy, and hence, the computation complexity reduces significantly. Moreover, the MFG formulation does not require explicit information exchange among the users or the users and the ES, which additionally requires $\mathcal{O}(N^2)$ (resp., $\mathcal{O}(N)$) amount of information exchange among the users (resp., between users and the ES). We will use this later to carry out extensive performance evaluation of the proposed MFG approach to demonstrate that the MFG approximates the N -user Nash game reasonably well.

This concludes the analysis of the offloading policy design for the MEC game with equitable access to all the users. In the next section, we will formulate the MEC system consisting of primary and secondary users and will compute NE policies.

V. MEC WITH PRIORITY-BASED ACCESS

Beginning with this section, we extend our earlier MEC framework (with equitable access) to an advanced setting comprised of one primary user and N secondary users as shown in Fig. 6. The primary user has priority access to the ES and the secondary users can offload computations to the ES by utilizing the transmitter of the primary user. The setting is deeply motivated by the *cognitive radio network* technology which aims to promote better effective utilization of the spectrum—a licensed resource. Within the same framework, the secondary users have limited access to the spectrum (which is originally reserved for the primary user) and are allowed to use it only when the primary user is inactive/idling. In this work, we propose a similar model for the MEC system⁴ wherein the primary user has priority access to a computing facility (which we also call the ES to maintain consistency with earlier sections) and secondary users can access the ES, but with a lower priority. We will formalize the priority discipline in a later section.

⁴Although we consider only one ES here, the techniques employed can be generalized to more complex settings with multiple ESs which are connected to the secondary users on a graph network. This setting lies within the realm of device-to-device/peer-to-peer communication in computer networks and oligopoly markets in economics and offers a promising future direction.

Our objective in this setup is again to solve for the NE policies for both the primary and the secondary users. However, the presence of the primary user leads to an interesting observation here compared to the setup of the previous sections. In the equitable access-MEC setup, all the users have a *vanishing* effect on the mean-field, i.e., the mean-field remains unaffected if finitely many users leave the system or deviate by behaving *irrationally*. In the current setup, however, the primary user (as we will also see later) has a non-vanishing effect. To handle this, we will leverage the paradigm of major-minor mean-field games (MM-MFGs) from control theory literature [12]. The idea is derived from banking systems where a *finite* number of “major” banks (or major players) can significantly affect the operations of “minor” banks (or minor players); however, the “minor” ones can only affect the operations of the “major” ones aggregatively (i.e., via their mean-field). With the above prelude to MM-MFGs, let us proceed with the formulation of the problem.

Consider a primary device D_P which can process the incoming tasks on either its local processor L_P or offload it to the ES using its transmitter T_P . Incoming tasks are distributed as a $\text{Poi}(\lambda_P)$ r.v. and are split for local processing or offloading using a Bernoulli r.v. with parameter p_P as shown in Fig. 6. The service time of the local processor and the transmitter of D_P are distributed as $\exp(\mu_{2P})$ and $\exp(P\mu_{1P})$ r.v.s., respectively, with the frequency parameter upper bounded as $\mu_{2P} \leq f_{P,max}$ and the power consumption parameter upper bounded as $\mu_{1P} \leq P_{max}$. The associated power consumed at the local processor is $\eta_P \mu_{2P}^3$ with $\eta_P > 0$.

Next, the parameters and service notions of each secondary user D_{S_i} and the ES are defined in the same way as for the MEC system of Section II, and hence, are not repeated here. We follow the *last-come-first-serve with priority preemption* (LCFS-PP) discipline, wherein if a primary device’s packet is being served at any server, any incoming secondary device’s packet is dropped and not allowed to be served. Only another packet of the primary user can preempt it. Secondary users however, get equitable access to both T_P and the ES and any secondary user can preempt the packet of any other secondary user. Both the primary and secondary users aim to minimize their power consumptions and average AoI of their offloaded packets. Additionally, in lieu of allowing a secondary user to access its transmitter, the primary user charges the secondary users an amount proportional to the loading created at T_P , which we will formalize next.

In contrast to the case of the MEC system with equitable access where there was no primary user, in this case, the transmitter of T_P is utilized by the secondary devices as well. Thus, the power expended for offloading is due to both its own packets and those of the secondary users whenever allowed. Thus, the fraction of time that the T_P is busy serving *its own packets* can be computed as

$$t_{T_P,1} = \lambda_P \bar{p}_P / (\lambda_P \bar{p}_P + \mu_{1P}). \quad (9)$$

Further, the fraction of time that the transmitter is busy serving packets of the secondary users can be computed as

$$t_{T_P,2} = \frac{\sum_{j=1}^N \lambda_j \bar{p}_j}{\sum_{j=1}^N \lambda_j \bar{p}_j + (\mu_{1P} + \lambda_P \bar{p}_P)} \frac{\mu_{1P}}{\mu_{1P} + \lambda_P \bar{p}_P}, \quad (10)$$

where the first multiplying fraction in the above expression denotes the average busy period of T_P serving the exogenous secondary input of $\lambda_s := \sum_{j=1}^N \lambda_j \bar{p}_j$ and the second one denotes the fraction of the time that the T_P is not serving the packets of the primary user. Subsequently, the mean service time spent by T_P in task offloading is obtained as $t_{T_P} = t_{T_P,1} + t_{T_P,2}$. In addition, the fraction of time that the primary device’s local processor is busy can be computed as

$$t_{L_P} = \lambda_P p_P / (\lambda_P p_P + \mu_{2P}). \quad (11)$$

Thus, we can formally state the primary and the secondary devices’ optimization problems as below.

Problem 3 (Primary device’s optimization problem):

$$\begin{aligned} \min_{(p_P, \mu_{1P}, \mu_{2P}) \in [0,1] \times \mathbb{R}^2} \quad & J_P^{(N)}(p_P, \mu_{1P}, \mu_{2P}, \lambda_s) \\ \text{s.t.} \quad & 0 \leq \mu_{1P} \leq P_{max} \\ & 0 \leq \mu_{2P} \leq f_{P,max}, \end{aligned}$$

where $J_P^{(N)}(p_P, \mu_{1P}, \mu_{2P}, \lambda_s) := t_{T_P} \mu_{1P} + t_{L_P} \eta_P \mu_{2P}^3 + V_P \Delta_P(p_P, \mu_{1P}, \mu_{2P}) - \alpha t_{T_P,2}$. Here, $\Delta_P(p_P, \mu_{1P}, \mu_{2P})$ denotes the average AoI incurred by the primary user and $\alpha \geq 0$ denotes the fixed price charged by the primary user for using its transmitter. Following this, the overall reward received $\alpha t_{T_P,2}$ from serving secondary users is proportional to the fraction of time that it serves secondary users.

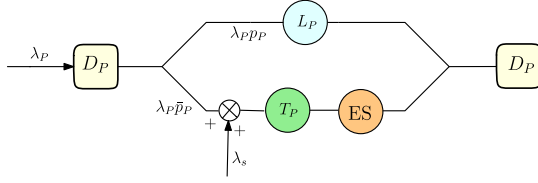
Next, let us define the joint decision variable vectors as $\mathbf{p}_{ps} := [p_1, \dots, p_N, p_P]$ and $\boldsymbol{\mu}_{2,ps} := [\mu_{21}, \dots, \mu_{2N}, \mu_{2P}]$, and the average AoI incurred by a secondary user as $\Delta_i^{(N)}(\mathbf{p}_{ps}, \boldsymbol{\mu}_{2,ps}, \mu_{1P})$. Then, the secondary device’s optimization problem is stated as follows.

Problem 4 (Secondary device’s optimization problem):

$$\begin{aligned} \min_{(p_i, \mu_{2i}) \in [0,1] \times \mathbb{R}} \quad & J_{S_i}^{(N)}(\mathbf{p}_{ps}, \boldsymbol{\mu}_{2,ps}, \mu_{1P}) \\ \text{s.t.} \quad & 0 \leq \mu_{2i} \leq f_{i,max}, \end{aligned}$$

where $J_{S_i}^{(N)}(\mathbf{p}_{ps}, \boldsymbol{\mu}_{2,ps}, \mu_{1P}) := t_{L_i} \eta_i \mu_{2i}^3 + V_i \Delta_i^{(N)}(\mathbf{p}_{ps}, \boldsymbol{\mu}_{2,ps}, \mu_{1P}) + \alpha t_{T_P,2} \lambda_i \bar{p}_i$, where we recall that $t_{L_i} = \lambda_i p_i / (\lambda_i p_i + \mu_{2i})$ denotes the fraction of time that the local processor L_{S_i} is busy. Further, $\alpha t_{T_P,2} \lambda_i \bar{p}_i$ denotes the *revenue* paid by the secondary user S_i to use the primary user’s transmitter and is a function of the fixed (per unit) price α , the service time fraction $t_{T_P,2}$ and the user’s own ‘mean transmission rate’ $\lambda_i \bar{p}_i$.

Now that we have formulated both the primary and secondary users’ optimization problems, we note again that the coupling between all the users due to the shared ES makes the underlying problem a non-cooperative game. Thus, for the same reasons as discussed earlier, we will leverage the paradigm of MM-MFGs to compute tractable equilibrium policies for solving the $(N+1)$ -user game problem. In this regard, we first need to characterize the average AoIs of both the primary and the secondary users which we do in the next section. Also, henceforth, we refer to (p_i, μ_{2i}) as the policy of secondary device D_{S_i} , $\forall i$, and $(p_P, \mu_{1P}, \mu_{2P})$ as the policy of primary device D_P .

Fig. 7. Task flow for primary device with exogenous incoming λ_s .TABLE IV
STATE DICTIONARY FOR THE FS-MC OF THE PRIMARY USER

state	server 1 (T_P)	server 2 (L_P)	server 3 (ES)
s_1	freshest	2^{nd} freshest	oldest
s_2	freshest	oldest	2^{nd} freshest
s_3	2^{nd} freshest	freshest	oldest
s_4	no packet	freshest	2^{nd} freshest
s_5	no packet	2^{nd} freshest	freshest

VI. AVERAGE AOI COMPUTATION

In this section, we will compute the average AoIs of the primary as well as the secondary users.

A. Average AoI Analysis for the Primary User

Let us begin by considering the perspective of the primary user as shown in Fig. 7, where we take the exogenous arrival process to be distributed as $\text{Poi}(\lambda_s)$ noting that $\lambda_s := \sum_{j=1}^N \lambda_j \bar{p}_j$, and all the incoming processes are Poisson distributed.

Next, we note that due to the priority access given to the primary device, on one hand, it can preempt the packet of any secondary device, and on the other hand, if a packet of the primary user is being served, the secondary user's task packet is dropped altogether. Thus, in essence, the secondary user does not have any effect on the average AoI of the primary user. Hence, for the purpose of computation, we can ignore the exogenous inputs due to the secondary users (i.e., set $\lambda_s = 0$ without loss of generality). Consequently, we can reduce the dimensionality of the FS-MC state space, which is elaborated upon as follows.

1) *State Space*: The state space S is comprised of 5 states which keep track of the server holding the freshest, the second freshest, and the oldest packets and/or is empty. Detailed descriptions are provided in Table IV. We employ fake updates in both L_P and the ES with the AoI of the fake packet equal to the AoI of the departing packet.

2) *Transition Functions*: Next, in Table V, we list the possible transitions in the FS-MC and the corresponding AoI vector $x'(t) := [x'_0(t) \ x'_1(t) \ x'_2(t) \ x'_3(t)]$. Let us take $\mathbf{u}_s := [1 \ 1 \ 1 \ 1]$, $\hat{\mathbf{u}}_s := [1 \ 0 \ 1 \ 1]$, $a := \lambda_P + \mu_{1P} + \mu_{2P} + \mu_3$ and $\hat{a} := a - \mu_{1P}$. Then, the steady state vector $\bar{\rho} := [\bar{\rho}_1, \bar{\rho}_2, \bar{\rho}_3, \bar{\rho}_4, \bar{\rho}_5]$ satisfies (2b) and the matrix equation:

$$A_P \bar{\rho} = M_P \bar{\rho}, \quad (12)$$

where the entries of the matrices A_P and M_P can be determined in a manner similar to (4), and are omitted due to space constraints (see detailed expressions in Appendix X-A).

TABLE V
STATE TRANSITIONS OF THE FS-MC AND ASSOCIATED AOI JUMPS FOR THE PRIMARY USER

s	$\mathbf{q}$	s'	$x' = xA_s$	$\mathbf{u}_s A_s$
s_1	$\lambda_P p_P$	s_3	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{10} \ \bar{v}_{11} \ 0 \ \bar{v}_{13}]$
	$\lambda_P \bar{p}_P$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{10} \ 0 \ \bar{v}_{12} \ \bar{v}_{13}]$
	μ_{1P}	s_5	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{10} \ 0 \ \bar{v}_{12} \ \bar{v}_{11}]$
	μ_{2P}	s_1	$[x_2 \ x_1 \ x_2 \ x_2]$	$[\bar{v}_{12} \ \bar{v}_{11} \ \bar{v}_{12} \ \bar{v}_{12}]$
	μ_3	s_1	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{13} \ \bar{v}_{11} \ \bar{v}_{12} \ \bar{v}_{13}]$
s_2	$\lambda_P p_P$	s_3	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{20} \ \bar{v}_{21} \ 0 \ \bar{v}_{23}]$
	$\lambda_P \bar{p}_P$	s_2	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{20} \ 0 \ \bar{v}_{22} \ \bar{v}_{23}]$
	μ_{1P}	s_5	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{20} \ 0 \ \bar{v}_{22} \ \bar{v}_{21}]$
	μ_{2P}	s_2	$[x_2 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{22} \ \bar{v}_{21} \ \bar{v}_{22} \ \bar{v}_{23}]$
	μ_3	s_2	$[x_3 \ x_1 \ x_3 \ x_3]$	$[\bar{v}_{23} \ \bar{v}_{21} \ \bar{v}_{23} \ \bar{v}_{23}]$
s_3	$\lambda_P p_P$	s_3	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{30} \ \bar{v}_{31} \ 0 \ \bar{v}_{33}]$
	$\lambda_P \bar{p}_P$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{30} \ 0 \ \bar{v}_{32} \ \bar{v}_{33}]$
	μ_{1P}	s_4	$[x_0 \ 0 \ x_2 \ x_1]$	$[\bar{v}_{30} \ 0 \ \bar{v}_{32} \ \bar{v}_{31}]$
	μ_{2P}	s_3	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{32} \ \bar{v}_{32} \ \bar{v}_{32} \ \bar{v}_{32}]$
	μ_3	s_3	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{33} \ \bar{v}_{31} \ \bar{v}_{32} \ \bar{v}_{33}]$
s_4	$\lambda_P p_P$	s_4	$[x_0 \ 0 \ 0 \ x_3]$	$[\bar{v}_{40} \ 0 \ 0 \ \bar{v}_{43}]$
	$\lambda_P \bar{p}_P$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{40} \ 0 \ \bar{v}_{42} \ \bar{v}_{43}]$
	μ_{2P}	s_4	$[x_2 \ 0 \ x_2 \ x_2]$	$[\bar{v}_{42} \ 0 \ \bar{v}_{42} \ \bar{v}_{42}]$
	μ_3	s_4	$[x_3 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{43} \ 0 \ \bar{v}_{42} \ \bar{v}_{43}]$
s_5	$\lambda_P p_P$	s_4	$[x_0 \ 0 \ 0 \ x_3]$	$[\bar{v}_{50} \ 0 \ 0 \ \bar{v}_{53}]$
	$\lambda_P \bar{p}_P$	s_2	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{50} \ 0 \ \bar{v}_{52} \ \bar{v}_{53}]$
	μ_{2P}	s_5	$[x_2 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{52} \ 0 \ \bar{v}_{52} \ \bar{v}_{53}]$
	μ_3	s_5	$[x_3 \ 0 \ x_3 \ x_3]$	$[\bar{v}_{53} \ 0 \ \bar{v}_{53} \ \bar{v}_{53}]$

Using the above set of linear equations in (12), one can compute the distribution $\bar{\rho}$ by combining with (2b). To compute the average AoI $\Delta_P(p_P, \mu_{1P}, \mu_{2P})$, it then remains to compute the steady state vector $\bar{\mathbf{v}} = [\bar{v}_1, \dots, \bar{v}_5]$ in Theorem 1 and then apply the formula for Δ in the theorem. To compute $\bar{\mathbf{v}}$, we use (3) in Theorem 1 to write down the set of linear equations satisfied by its components, which can be compactly written using the linear matrix equation

$$A_P \bar{\mathbf{v}} = V_P \bar{\mathbf{v}}. \quad (14)$$

The entries of the matrix V_P can be determined similarly to (5) and are omitted due to space constraints (see Appendix X-A for details). Consequently, the average AoI can be computed by first computing $\bar{\rho}$ using (12), substituting the same in (14), and subsequently solving the latter. We can then summarize the above result in the following theorem.

Theorem 3: Suppose that the arrival rate at the primary device is distributed as $\text{Poi}(\lambda_P)$ and the service rates as $\exp(\mu_{1P})$ and $\exp(\mu_{2P})$ at the transmitter T_P and the local processor L_P , respectively. Let the service rate of the ES be distributed as $\exp(\mu_3)$. Then, the average AoI of the primary user $\Delta_P(p_P, \mu_{1P}, \mu_{2P})$ is given in (13a) shown at the bottom of the next page, and is obtained by solving the sets of equations in (12) and (14).

The AoI computation for the primary user is now complete. Next, we proceed toward computing the average AoI for the secondary users.

B. Average AoI Analysis for the Secondary User

Let us now consider the perspective of a secondary user S_i as shown in Fig. 8 with two types of exogeneous incomings, (1) a

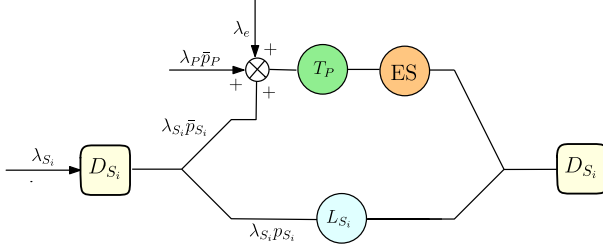


Fig. 8. Task flow for secondary device S_i with exogeneous incoming λ_e and $\lambda_P \bar{p}_P$.

TABLE VI
STATE DICTIONARY FOR FS-MC OF THE SECONDARY USER S_i

state	server 1 (T_P)	server 2 (L_{S_i})	server 3 (ES)
s_1	freshest	2^{nd} freshest	oldest
s_2	freshest	oldest	2^{nd} freshest
s_3	freshest	2^{nd} freshest	class P
s_4	2^{nd} freshest	freshest	oldest
s_5	class P	freshest	2^{nd} freshest
s_6	2^{nd} freshest	freshest	class P
s_7	class P	freshest	class P
s_8	class P	2^{nd} freshest	freshest
s_9	oldest	freshest	2^{nd} freshest
s_{10}	oldest	2^{nd} freshest	freshest

Poisson process with parameter $\lambda_P \bar{p}_P$ due to the primary user, and (2) a Poisson process with parameter $\lambda_e := \sum_{j \in [N], j \neq i} \lambda_j \bar{p}_j$ due to the other secondary users. The state space and the transition probabilities of the FS-MC for this system in Fig. 8 are given as follows.

1) *State Space*: The state space S is comprised of 10 states which keep track of the server holding the freshest, the second freshest, and the oldest packets of S_i , $\forall i$ and whether a server holds a packet from the primary user or not. We will refer to the packet of the primary user as ‘class P’ packet. Detailed descriptions are provided in Table VI. Furthermore, we run fake updates in all the servers where the fake packet has low priority with its AoI being the age of the departing packet.

2) *Transition functions*: Next, in Table VII, we list the possible transitions in the FS-MC and the corresponding AoI vector $x'(t) := [x'_0(t) \ x'_1(t) \ x'_2(t) \ x'_3(t)]$, where x'_0, x'_1, x'_2 , and x'_3 denote the AoIs of the packets at the device D_{S_i} , T_P , L_{S_i} , and the ES, respectively. Henceforth, we will suppress the index i for brevity. Similar to the earlier analysis, let us take $\mathbf{u}_s := [1 \ 1 \ 1 \ 1]$ and $a := \lambda + \lambda_e + \mu_1 + \mu_2 + \mu_3$. Then, the steady state vector $\bar{\pi} = [\bar{\pi}_1, \dots, \bar{\pi}_{10}]$ for the secondary user S_i satisfies (2b) and

the linear matrix equation:

$$A_S \bar{\pi} = M_S \bar{\pi}, \quad (15)$$

where the entries of matrices A_S and M_S are omitted due to space constraints (see Appendix X-B for details). Subsequently, using (3), we can use the solution to (15) to write down the equations satisfied by the steady-state conditional distribution vector $\bar{v} = [\bar{v}_1, \dots, \bar{v}_{10}]$ for the secondary user S_i in a compact matrix equation as

$$A_S \bar{v} = V_S \bar{v}, \quad (16)$$

the entries of the matrix V_S are provided in Appendix X-B. Then, we solve this linear system of equations to obtain the average AoI $\Delta_i^{(N)}(\mathbf{p}_{ps}, \mu_{2,ps}, \mu_{1P})$ which we summarize in the next theorem with the use of notation i for indexing.

Theorem 4: Suppose that the arrival rate at a secondary device S_i is distributed as $\text{Poi}(\lambda_{S_i})$ and the service rate of its local processor as $\exp(\mu_{2i})$. Further, let the service rates of the transmitter of the primary user and the ES be distributed as $\exp(\mu_{1P})$ and $\exp(\mu_3)$, respectively. Then, the average AoI $\Delta_i^{(N)}(\mathbf{p}_{ps}, \mu_{2,ps}, \mu_{1P})$ of S_i exists for all i and is obtained by solving the sets of equations in (15) and (16).

Theorem 3 and Theorem 4 completely characterize the average AoI of the primary and the secondary users, respectively, and hence, also Problems 3 and 4. Due to long expressions of $\Delta_i^{(N)}(\mathbf{p}_{ps}, \mu_{2,ps}, \mu_{1P})$, we do not provide them here; however, later, we will provide an algorithm to compute the equilibrium policies without requiring their explicit forms. We are now ready to provide equilibrium policies for the MEC system with priority-based access.

VII. MM-MF ANALYSIS FOR THE MEC SYSTEM

Our aim here is again to compute Nash equilibrium policies for the primary and the secondary users. However, for reasons of tractability discussed before, we appeal to the MM-MFG framework to compute completely *distributed* approximate NE solutions using only the local information of the users.

To this end, let us define, using the same notations as in previous sections, the quantities $\rho^{(N)} := \frac{\lambda_e}{N\mu_{1P}}$ and $\rho := \lim_{N \rightarrow \infty} \rho^{(N)}$, where $\rho^{(N)}$ denotes the mean load on the primary device’s transmitter T_P and ρ denotes the MF approximation as discussed earlier. Then, we have that for a large user MEC system, the exogenous arrival rates λ_e and λ_s can be approximated as

$$\lambda_e = (N - 1)\mu_{1P} \times \lambda_e / (N - 1)\mu_{1P} \approx (N - 1)\mu_{1P}\rho, \quad (17)$$

$$\lambda_s = N\mu_{1P} \times \lambda_s / N\mu_{1P} \approx N\mu_{1P}\rho. \quad (18)$$

$$\begin{aligned} \Delta_P(p_P, \mu_{1P}, \mu_{2P}) = & \frac{\mu_{1P} + \mu_{2P} + \mu_3}{(\mu_{1P} + \mu_{2P})(\mu_{2P} + \mu_3)} + \frac{\mu_{1P}\mu_3(\lambda_P + \mu_{2P})}{\lambda_P(\mu_{1P} + \mu_{2P})(\mu_{2P} + \mu_3)(\lambda_P \times \bar{p}_P + \mu_{2P})} \\ & - \frac{\mu_{2P}\mu_3(\lambda_P + \mu_{1P})}{\lambda_P(\mu_{1P} + \mu_{2P})(\mu_{1P} - \mu_3)(\mu_{1P} + \lambda_P \times p_P)} + \frac{\mu_{1P}\mu_{2P}(\lambda_P + \mu_3)}{\lambda_P(\mu_{1P} - \mu_3)(\mu_{2P} + \mu_3)(\mu_3 + \lambda_P \times p_P)} \end{aligned} \quad (13a)$$

TABLE VII
STATE TRANSITIONS OF THE FS-MC AND ASSOCIATED AOI JUMPS FOR THE SECONDARY USER

s	q	s'	$x' = xA_s$	$v_s A_s$
s_1	λ_P	s_4	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{10} \ \bar{v}_{11} \ 0 \ \bar{v}_{13}]$
	$\lambda_{\bar{P}}$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{10} \ 0 \ \bar{v}_{12} \ \bar{v}_{13}]$
	λ_e	s_9	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{10} \ \bar{v}_{10} \ \bar{v}_{12} \ \bar{v}_{13}]$
	λ_P	s_5	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{10} \ \bar{v}_{10} \ \bar{v}_{12} \ \bar{v}_{13}]$
	μ_1	s_{10}	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{10} \ \bar{v}_{11} \ \bar{v}_{12} \ \bar{v}_{11}]$
	μ_2	s_1	$[x_2 \ x_1 \ x_2 \ x_2]$	$[\bar{v}_{12} \ \bar{v}_{11} \ \bar{v}_{12} \ \bar{v}_{12}]$
s_2	λ_P	s_4	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{20} \ \bar{v}_{21} \ 0 \ \bar{v}_{23}]$
	$\lambda_{\bar{P}}$	s_2	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{20} \ 0 \ \bar{v}_{22} \ \bar{v}_{23}]$
	λ_e	s_{10}	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{20} \ \bar{v}_{20} \ \bar{v}_{22} \ \bar{v}_{23}]$
	λ_P	s_8	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{20} \ \bar{v}_{20} \ \bar{v}_{22} \ \bar{v}_{23}]$
	μ_1	s_{10}	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{20} \ \bar{v}_{21} \ \bar{v}_{22} \ \bar{v}_{21}]$
	μ_2	s_2	$[x_2 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{22} \ \bar{v}_{21} \ \bar{v}_{22} \ \bar{v}_{23}]$
s_3	λ_P	s_6	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{30} \ \bar{v}_{31} \ 0 \ \bar{v}_{33}]$
	$\lambda_{\bar{P}}$	s_3	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{30} \ 0 \ \bar{v}_{32} \ \bar{v}_{33}]$
	λ_e	s_6	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{30} \ \bar{v}_{30} \ \bar{v}_{32} \ \bar{v}_{33}]$
	λ_P	s_7	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{30} \ \bar{v}_{30} \ \bar{v}_{32} \ \bar{v}_{33}]$
	μ_1	s_6	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{30} \ \bar{v}_{30} \ \bar{v}_{32} \ \bar{v}_{33}]$
	μ_2	s_3	$[x_2 \ x_1 \ x_2 \ x_2]$	$[\bar{v}_{32} \ \bar{v}_{31} \ \bar{v}_{32} \ \bar{v}_{32}]$
s_4	λ_P	s_6	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{40} \ \bar{v}_{41} \ 0 \ \bar{v}_{43}]$
	$\lambda_{\bar{P}}$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{40} \ 0 \ \bar{v}_{42} \ \bar{v}_{43}]$
	λ_e	s_9	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{40} \ \bar{v}_{40} \ \bar{v}_{42} \ \bar{v}_{43}]$
	λ_P	s_5	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{40} \ \bar{v}_{40} \ \bar{v}_{42} \ \bar{v}_{43}]$
	μ_1	s_9	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{40} \ \bar{v}_{41} \ \bar{v}_{42} \ \bar{v}_{41}]$
	μ_2	s_4	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{42} \ \bar{v}_{42} \ \bar{v}_{42} \ \bar{v}_{42}]$
s_5	λ_P	s_4	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{43} \ \bar{v}_{41} \ \bar{v}_{42} \ \bar{v}_{43}]$
	λ_P	s_5	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{50} \ \bar{v}_{51} \ 0 \ \bar{v}_{53}]$
	$\lambda_{\bar{P}}$	s_5	$[x_0 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{50} \ \bar{v}_{51} \ \bar{v}_{52} \ \bar{v}_{53}]$
	λ_e	s_5	$[x_0 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{50} \ \bar{v}_{51} \ \bar{v}_{52} \ \bar{v}_{53}]$
	λ_P	s_5	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{50} \ \bar{v}_{50} \ \bar{v}_{52} \ \bar{v}_{53}]$
	μ_1	s_6	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{50} \ \bar{v}_{51} \ \bar{v}_{52} \ \bar{v}_{51}]$
s_6	λ_P	s_6	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{60} \ \bar{v}_{61} \ 0 \ \bar{v}_{63}]$
	$\lambda_{\bar{P}}$	s_3	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{60} \ 0 \ \bar{v}_{62} \ \bar{v}_{63}]$
	λ_e	s_6	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{60} \ \bar{v}_{60} \ \bar{v}_{62} \ \bar{v}_{63}]$
	λ_P	s_7	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{60} \ \bar{v}_{60} \ \bar{v}_{62} \ \bar{v}_{63}]$
	μ_1	s_6	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{60} \ \bar{v}_{60} \ \bar{v}_{62} \ \bar{v}_{63}]$
	μ_2	s_6	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{62} \ \bar{v}_{62} \ \bar{v}_{62} \ \bar{v}_{62}]$
s_7	λ_P	s_4	$[x_3 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{63} \ \bar{v}_{61} \ \bar{v}_{62} \ \bar{v}_{63}]$
	λ_P	s_7	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{70} \ \bar{v}_{71} \ 0 \ \bar{v}_{73}]$
	$\lambda_{\bar{P}}$	s_7	$[x_0 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{70} \ \bar{v}_{71} \ \bar{v}_{72} \ \bar{v}_{73}]$
	λ_e	s_7	$[x_0 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{70} \ \bar{v}_{71} \ \bar{v}_{72} \ \bar{v}_{73}]$
	λ_P	s_7	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{70} \ \bar{v}_{70} \ \bar{v}_{72} \ \bar{v}_{73}]$
	μ_1	s_6	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{70} \ \bar{v}_{71} \ \bar{v}_{72} \ \bar{v}_{71}]$
s_8	λ_P	s_7	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{72} \ \bar{v}_{72} \ \bar{v}_{72} \ \bar{v}_{72}]$
	μ_2	s_7	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{72} \ \bar{v}_{72} \ \bar{v}_{72} \ \bar{v}_{72}]$
	μ_3	s_5	$[x_3 \ x_3 \ x_3 \ x_3]$	$[\bar{v}_{73} \ \bar{v}_{73} \ \bar{v}_{72} \ \bar{v}_{73}]$
	λ_P	s_5	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{80} \ \bar{v}_{81} \ 0 \ \bar{v}_{83}]$
	$\lambda_{\bar{P}}$	s_8	$[x_0 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{80} \ \bar{v}_{81} \ \bar{v}_{82} \ \bar{v}_{83}]$
	λ_e	s_8	$[x_0 \ x_1 \ x_2 \ x_3]$	$[\bar{v}_{80} \ \bar{v}_{81} \ \bar{v}_{82} \ \bar{v}_{83}]$
s_9	λ_P	s_8	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{80} \ \bar{v}_{80} \ \bar{v}_{82} \ \bar{v}_{83}]$
	μ_1	s_6	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{80} \ \bar{v}_{81} \ \bar{v}_{82} \ \bar{v}_{81}]$
	μ_2	s_8	$[x_2 \ x_2 \ x_2 \ x_3]$	$[\bar{v}_{82} \ \bar{v}_{82} \ \bar{v}_{82} \ \bar{v}_{83}]$
	μ_3	s_8	$[x_3 \ x_3 \ x_3 \ x_3]$	$[\bar{v}_{83} \ \bar{v}_{83} \ \bar{v}_{83} \ \bar{v}_{83}]$
	λ_P	s_9	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{90} \ \bar{v}_{91} \ 0 \ \bar{v}_{93}]$
	$\lambda_{\bar{P}}$	s_1	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{90} \ 0 \ \bar{v}_{92} \ \bar{v}_{93}]$
s_{10}	λ_e	s_9	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{90} \ \bar{v}_{90} \ \bar{v}_{92} \ \bar{v}_{93}]$
	λ_P	s_5	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{90} \ \bar{v}_{90} \ \bar{v}_{92} \ \bar{v}_{93}]$
	μ_1	s_9	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{90} \ \bar{v}_{91} \ \bar{v}_{92} \ \bar{v}_{91}]$
	μ_2	s_9	$[x_2 \ x_2 \ x_2 \ x_2]$	$[\bar{v}_{92} \ \bar{v}_{92} \ \bar{v}_{92} \ \bar{v}_{92}]$
	μ_3	s_9	$[x_3 \ x_3 \ x_2 \ x_3]$	$[\bar{v}_{93} \ \bar{v}_{93} \ \bar{v}_{92} \ \bar{v}_{93}]$
	λ_P	s_9	$[x_0 \ x_1 \ 0 \ x_3]$	$[\bar{v}_{10,0} \ \bar{v}_{10,1} \ 0 \ \bar{v}_{10,3}]$
s_{10}	$\lambda_{\bar{P}}$	s_2	$[x_0 \ 0 \ x_2 \ x_3]$	$[\bar{v}_{10,0} \ 0 \ \bar{v}_{10,2} \ \bar{v}_{10,3}]$
	λ_e	s_{10}	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{10,0} \ \bar{v}_{10,0} \ \bar{v}_{10,2} \ \bar{v}_{10,3}]$
	λ_P	s_8	$[x_0 \ x_0 \ x_2 \ x_3]$	$[\bar{v}_{10,0} \ \bar{v}_{10,0} \ \bar{v}_{10,2} \ \bar{v}_{10,3}]$
	μ_1	s_9	$[x_0 \ x_1 \ x_2 \ x_1]$	$[\bar{v}_{10,0} \ \bar{v}_{10,1} \ \bar{v}_{10,2} \ \bar{v}_{10,1}]$
	μ_2	s_{10}	$[x_2 \ x_2 \ x_2 \ x_3]$	$[\bar{v}_{10,2} \ \bar{v}_{10,2} \ \bar{v}_{10,2} \ \bar{v}_{10,3}]$
	μ_3	s_{10}	$[x_3 \ x_3 \ x_3 \ x_3]$	$[\bar{v}_{10,3} \ \bar{v}_{10,3} \ \bar{v}_{10,3} \ \bar{v}_{10,3}]$

Consequently, the primary user's optimization problem is as stated below.

Problem 5 (Primary device's MF optimization problem):

$$\begin{aligned}
 \min_{(p_P, \mu_{1P}, \mu_{2P}) \in [0,1] \times \mathbb{R}^2} \quad & J_{P,\rho}(p_P, \mu_{1P}, \mu_{2P}) \\
 \text{s.t.} \quad & 0 \leq \mu_{1P} \leq P_{max} \\
 & 0 \leq \mu_{2P} \leq f_{P,max},
 \end{aligned}$$

where $J_{P,\rho}(p_P, \mu_{1P}, \mu_{2P}) := J_P^{(N)}(p_P, \mu_{1P}, \mu_{2P}, \lambda_s)$ $|\lambda_s = N\mu_{1P}\rho$ denotes the cost function of the primary user as a function of its decision variables and the mean-field generated by the "minor" secondary users.

Next, for a generic secondary user of type ϕ with packet arrivals distributed as $\text{Poi}(\lambda_\phi)$, we split the incomings using a mean p_ϕ i.i.d. Bernoulli distributed r.v. and the service time of the local processor distributed as $\exp(\mu_{2,\phi})$ with $\mu_{2,\phi} \leq f_{\phi,max}$. Thus, we can state the generic secondary device's local optimization problem as follows.

Problem 6 (Secondary device MF optimization problem):

$$\min_{(p_\phi, \mu_{2,\phi}) \in [0,1] \times \mathbb{R}} \quad J_{S_\phi,\rho}(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}, p_P, \mu_{1P}, \mu_{2P})$$

$$\text{s.t.} \quad 0 \leq \mu_{2,\phi} \leq f_{\phi,max},$$

where $J_{S_\phi,\rho}(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}, p_P, \mu_{1P}, \mu_{2P}) := J_{S_i}^{(N)}(p_{ps}, \mu_{2,ps}, \mu_{1P}) |_{\lambda_e = (N-1)\mu_{1P}\rho, \lambda_s = N\mu_{1P}\rho}$.

Consequently, analogous to the equitable-access MEC case, the MFG is defined using three operators, namely, the optimality operators for the primary and secondary user of type ϕ , and the consistency operators as follows:

1) Optimality for the primary user:

$$(\hat{p}_P, \hat{\mu}_{1P}, \hat{\mu}_{2P}) = \bar{\Psi}_1(\rho) := \operatorname{argmin} J_{P,\rho}(p_P, \mu_{1P}, \mu_{2P})$$

subject to the constraints in Problem 5.

2) Optimality for the secondary user of type $\phi, \forall \phi \in \Phi$:

$$\begin{aligned}
 (\hat{p}_\phi, \hat{\mu}_{2,\phi}) &= \bar{\Psi}_2(\rho) \\
 &:= \operatorname{argmin} J_{S_\phi,\rho}(p_\phi, \mu_{1,\phi}, \mu_{2,\phi}, \hat{p}_P, \hat{\mu}_{1P}, \hat{\mu}_{2P})
 \end{aligned}$$

subject to the constraints in Problem 6.

3) Consistency: $\hat{\rho} = \bar{\Psi}_3(\hat{p}_\phi, \hat{\mu}_{1,\phi}, \hat{\mu}_{2,\phi}) := \mathbb{E}_{\mathbb{P}(\phi)}[\frac{\lambda_\phi \hat{p}_\phi}{\mu_{1P}}]$.

Briefly, the optimality operator $\bar{\Psi}_1(\cdot)$ outputs an optimal policy $(\hat{p}_P, \hat{\mu}_{1P}, \hat{\mu}_{2P})$ for the primary user for a given MF ρ . Subsequently, the optimality operator $\bar{\Psi}_2(\cdot)$ outputs an optimal policy $(\hat{p}_\phi, \hat{\mu}_{2,\phi})$ for a secondary user of type ϕ for the same

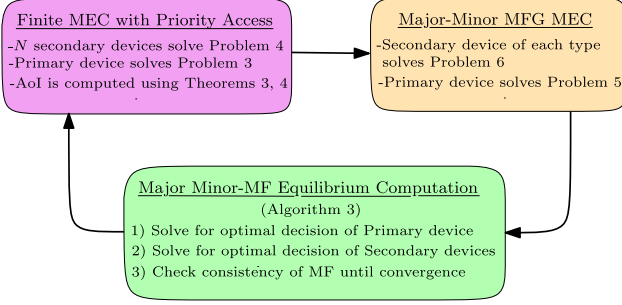


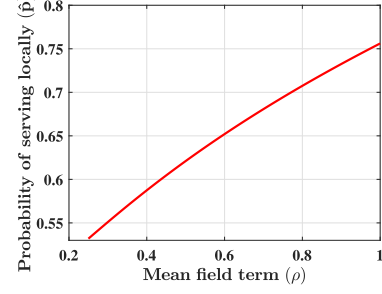
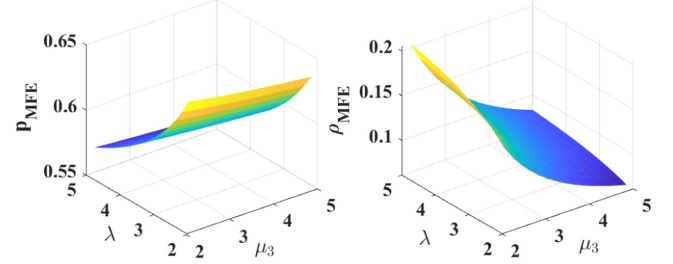
Fig. 9. Solution approach for the MEC setup with priority access.

Algorithm 2: Fixed-Point Iteration for Computing an MM-MFE Policy.

- 1: **Input:** $V_\phi, V_P, \eta_\phi, \eta_P, \lambda_\phi, \lambda_P, \mu_3, \forall \phi$ # MEC parameters
 - 2: **Input:** $\varepsilon_1, \varepsilon_2, \varepsilon_3$ # tolerance parameters
 - 3: **Input:** $\gamma_1, \gamma_2, \gamma_3$ # Iteration step sizes
 - 4: **Initialize:** $\hat{\rho}^{(0)}, \sigma_\phi^{(0)} := (p_\phi^{(0)}, \mu_{1,\phi}^{(0)}, \mu_{2,\phi}^{(0)}), \forall \phi$
 - 5: **Initialize:** $\sigma_P^{(0)} := (p_P^{(0)}, \mu_{1P}^{(0)}, \mu_{2P}^{(0)})$
 - 6: **while** $|\hat{\rho}^{(m)} - \hat{\rho}^{(m-1)}| < \varepsilon_1$ **do**
 - 7: Compute the optimal policy for the primary user using lines 7-18 of Algorithm 1 and Problem 5
 - 8: Given the primary user's policy, compute the optimal policy for secondary users using lines 6-19 of Algorithm 1 and Problem 6.
 - 9: $\hat{\rho}^{(k)} \leftarrow (1 - \gamma_1)\hat{\rho}^{(k-1)} + \gamma_1 \mathbb{E}_P(\phi) \left[\frac{\lambda_\phi^{(k)}(1 - \hat{p}_\phi^{(k)})}{\mu_{1P}^{(k)}} \right]$
 - 10: **end while**
 - 11: **Output:** $\hat{\rho}^{(m)}, \sigma_P^{(m)}, \sigma_\phi^{(m)}, \forall \phi$.
-

value of ρ and the obtained values of $(\hat{p}_P, \hat{\mu}_{1P}, \hat{\mu}_{2P})$ from $\bar{\Psi}_1(\cdot)$. Finally, the consistency operator $\bar{\Psi}_3(\cdot)$ generates a new ρ by using the optimal policies obtained above using $\bar{\Psi}_1$ and $\bar{\Psi}_2$, and signifies the consistent behavior of the population of the secondary users. The MFE which constitutes the tuple of equilibrium policies of the primary user, $(p_{P,MFE}, \mu_{1P,MFE}, \mu_{2P,MFE})$, the secondary users for all types, $(p_{\phi,MFE}, \mu_{1,\phi,MFE}, \mu_{2,\phi,MFE})_{\forall \phi}$, and the equilibrium mean-field (ρ_{MFE}) , is given by the fixed point of the composite map of Ψ_1, Ψ_2 , and Ψ_3 . We provide Algorithm 2 to compute the MM-MFE. Briefly, for a given value of ρ , we first solve Problem 5 to compute an optimal policy for the primary user (line 7 of Algorithm 2). Then, given this policy, we solve Problem 6 for the secondary users (line 8 of Algorithm 2). Finally, we update the mean-field using the consistency condition (line 9 of Algorithm 2). The MFE is then given by the fixed point of the composition of the maps Ψ_1, Ψ_2 , and Ψ_3 . A flowchart of the MM-MFG solution approach is presented in Fig. 9 for a better understanding.

Now, we have completely defined the MEC problem with priority access using MM-MFGs. In the next section, we will provide simulations to extensively evaluate the performance of the proposed approach for both the MEC with equitable access case and the MEC with priority access.

Fig. 10. The optimal probability $\hat{p}$ as a function of the MF term ρ .Fig. 11. Variation of the equilibrium probability p_{MFE} and equilibrium load ρ_{MFE} with generic arrival rate λ and service rate μ_3 .

VIII. PERFORMANCE EVALUATION

In this section, we perform a rigorous performance evaluation of our proposed MFG approach to design optimal offloading policies for users. Since this is the first work employing the mean-field games formulation, a direct comparative study with literature may not be possible at this time; thus we perform an extensive validation and refer to historically observed trends, wherever applicable.

A. The MEC System With Equitable Access

In this subsection, we provide numerical analysis for the MEC system with equitable access. We consider a homogeneous population for the first 4 studies and heterogeneous for the 5th study, for ease in illustration of the main concepts.

1) *Variation of p vs ρ :* In the first numerical study of this subsection, we plot in Fig. 10, the variation of local server usage versus the ES loading, denoted by the mean-field term ρ . We take the parameters to be $V = 10, \eta = 0.5, \lambda_\phi = 1, P_{\phi,max} = 1$, and $f_{\phi,max} = 0.8$. From Fig. 10, we observe that as the mean load at the ES increases (on the x -axis), the optimal probability of using the local processor (on the y -axis) increases and that of offloading to the ES decreases. This should be expected since the devices care about the AoI $V = 10$ times more than the average power consumed, as depicted by the value of V . Thus, if the ES is heavily loaded, the device is better-off serving tasks locally to incur a lower AoI.

2) *Effect of λ and μ_3 on MFE:* Next, in Fig. 11, we plot the variation of MFE as a function of the arrival rate λ and the service rate of the ES μ_3 . We consider the same set of parameters as in the previous study. In Fig. 11, we observe that the equilibrium loading at the ES decreases (as seen from the right subplot) as μ_3 increases with the decreasing probability of processing

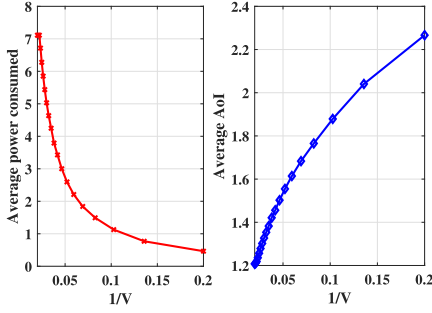


Fig. 12. The average power and average AoI per user at equilibrium as a function of the inverse weighting parameter $1/V$ for $N = 60$.

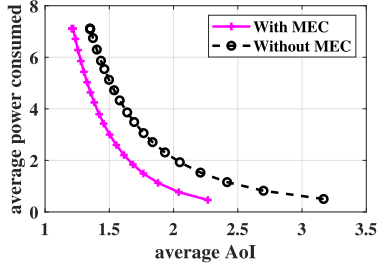


Fig. 13. Pareto front for a device w/ and w/o the presence of the ES.

locally. In other words, as we increase μ_3 , the total number of packets processed at the edge server increases. However, since the service rate μ_3 increases, the load on the edge server ρ_{MFE} still decreases. The decrease in ρ_{MFE} is low due to the fact that high V restrains users from incurring too much staleness as a result of offloading. Furthermore, the equilibrium offloading probability $1 - p_{\text{MFE}}$ increases with increasing arrival rate for given ES service rate μ_3 to accommodate the higher arrival rate, which consequently also increases ρ_{MFE} .

3) *Effect of V on the power consumed and AoI incurred:* Next, in Fig. 12, we study the relationship between the average power consumption/average AoI incurred by the devices at equilibrium and the inverse weighting parameter $1/V$. We take the parameters to be $\eta = 15$, $\lambda = 10$, $\mu_3 = 15$, $P_{\phi, \max} = 1$, $f_{\phi, \max} = 0.8$, and $N = 60$. We see from the left plot in Fig. 12 that the consumed power varies inversely with $1/V$ and converges to the value 0.464. Meanwhile, the average AoI incurred increases with $1/V$, as seen from the right plot. We would like to mention a couple of noteworthy points here. First, the trend observed through these plots closely resembles the $[O(V), O(1/V)]$ relationship observed in the works [13], [15]. A direct comparison between our work and theirs, however, is not fair since the nature of optimization differs (static in our case and dynamic in theirs) and the solution approaches used are widely contrasting (distributed in our case and centralized in theirs). Nevertheless, it is worth noting that the resource allocation method used in their work causes an increase in the time complexity of computation with the number of users, and thus simulating a large number of agents can be challenging (as seen from their simulation results). In contrast, our MFG approach allows us to compute low complexity solutions, which can be scaled to a possibly large-user population. Additionally, from Fig. 13, we observe

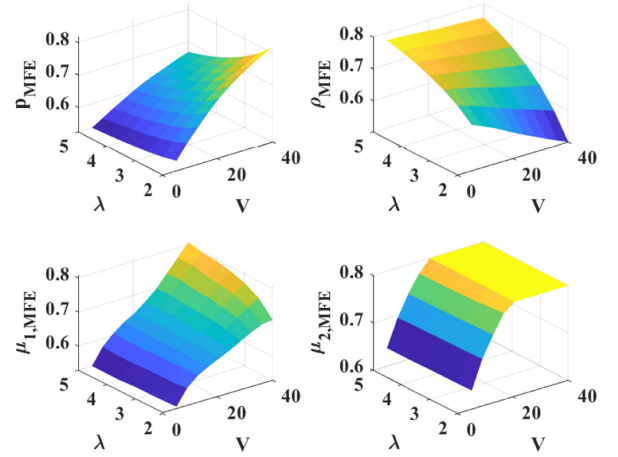


Fig. 14. Variation of the equilibrium policy and equilibrium load ρ_{MFE} with generic arrival rate λ and weighting constant V .

that the Pareto front of the solution with an MEC system lies below the one without the provision of an ES (no MEC), which demonstrates the significance of edge computing. The same observation is also noted in detail in the next study. On the other hand, if a completely greedy offloading policy is pursued, i.e., $p_{\phi} = 0$ and $\mu_{1\phi} = 1, \forall \phi$, then the average AoI incurred is 44.139 while the average power consumed is 0.9091, which leads to a significantly higher cost as compared to the solution obtained using our approach (AoI 1.879, power 1.127), indicating that such a greedy approach is suboptimal.

4) *Effect of λ and V on MFE:* Next, in Fig. 14, we study the dependence of the MFE solution on the arrival rates and the weighting constant for the same set of parameters as for the above study. We first observe that higher weighting on the AoI incentivizes the devices to process locally thereby increasing p_{MFE} . However, p_{MFE} increases only slightly with increasing λ . This is due to the fact that the local processor saturates at its maximum service rate after which more preemptions start causing an increase in the AoI, and hence it is suggested that we increase the transmissions to the ES as well. The latter phenomenon is apparent from the top right plot which shows that the ES loading increases for increasing λ for a given V . Also, from the top right plot, we observe that for high arrival rates, even though V increases, the ES equilibrium load increases. This suggests that once the local processor saturates, there is no option left for the device but to offload to the ES to manage the high rate incoming tasks. Further, we can also observe the local processor saturation from the bottom right plot with increase in V and λ at its maximum operating frequency $\mu_{1, \text{MFE}}$ at equilibrium. This then causes a simultaneous increase in the transmission of packets, which can be seen by an increase in the transmission power $\mu_{1, \text{MFE}}$ of the transmitter, to compensate for the increase in average AoI due to saturation. This observation also helps us demonstrate the necessity of an MEC system, and verify the “quality of experience” of computation in an MEC system since without an ES, the AoI would keep increasing due to local saturation thereby leading to a computation bottleneck at the device.

TABLE VIII

THE FIRST TABLE SHOWS THE COMPARISON OF A LOCAL NE POLICY WITH THE LOCAL MFE POLICY; THE SECOND TABLE SHOWS THE COMPARISON OF THE COST PER USER UNDER THE TWO POLICIES

N	$(\mu_{1,NE}, \mu_{2,NE}, p_{NE})$	$(\mu_{1,MFE}, \mu_{2,MFE}, p_{MFE})$
30	(0.5603, 0.6007, 0.5202)	(0.5647, 0.6474, 0.5701)
40	(0.5644, 0.6117, 0.5233)	(0.5452, 0.6196, 0.5681)
50	(0.5719, 0.6363, 0.5314)	(0.5322, 0.5969, 0.5655)

N	cost under NE policy	cost under MFE policy
30	16.0739	13.5142
40	21.5955	18.9599
50	26.4713	24.6980

5) *Comparison of NE and MFE policies:* Finally, we compare the performance of the proposed MFE approach (computed using Algorithm 1) with the corresponding Nash equilibrium (NE) performance (computed using Algorithm 3 in the Appendix). The users are uniformly distributed into 5 types, each with different values of $\lambda_\phi \in \{2, 3, 4, 5, 6\}$, $f_{\phi,max} \in \{0.8, 0.9, 1, 1.3, 1.6\}$, and $P_{\phi,max} \in \{1, 1.1, 1.2, 1.3, 1.6\}$ and with $V = 10$ and $\eta = 0.5$ as in the earlier studies. The results of the values obtained are tabulated in Table VIII for three different values of $N = 30, 40$, and 50 . Here, we observe that the MFE policy approximates the NE policy reasonably well in the sense that the both of them yield nearly the same performance for each user.

Our numerical study of the equitable access MEC game is now complete. Next, we provide numerical analysis for the MEC system comprised of one primary and multiple secondary users and numerically compute equilibrium policies.

B. The MEC With Priority Access

1) *Effect of α on MFE and associated costs:* In the first study of this subsection, we plot Fig. 15 which shows the effect of variations of the pricing parameter α on different quantities in the MM-MEC MFG. We take $N = 30$, $f_{P,max} = 3$, $f_{\phi,max} = 1.5$, $V = 10$, $\eta = 0.5$, $P_{max} = 10$, $\lambda_P = 4$, $\lambda_\phi = 5$, and $\mu_3 = 15$. The initial conditions for the primary user were set to $(\mu_{1P}, \mu_{2P}, p_P) = (0.6, 0.3, 0.5)$ and those for the secondary users to be $(\mu_{2,\phi}, p_\phi) = (0.6, 0.5)$. The quantities relating to the primary user are plotted in pink color, the ones related to the secondary user are plotted in blue and the mean load is plotted in red color. First, we observe that, with increasing price per unit, the equilibrium cost J_P^* of the primary user decreases due to increasing revenue earned. Meanwhile, the local processing cost for the secondary user $J_{L,\phi}^*$ increases since with a higher price charged by the primary user, the secondary user prefers to use its local processor more. This is further confirmed by its equilibrium policy constituting its local processing frequency $\mu_{2,\phi,MFE}$ and its probability of serving locally $p_{\phi,MFE}$, both of which also increase. The decrease in the total cost J_ϕ^* of the secondary user is due to a better average AoI as a consequence of fewer offloads, which counteracts the increase in the price charged by the primary user. The equilibrium policy $(\mu_{1P,MFE}, \mu_{2P,MFE}, p_{P,MFE})$ shows only a slight change for different values of α . This is due to

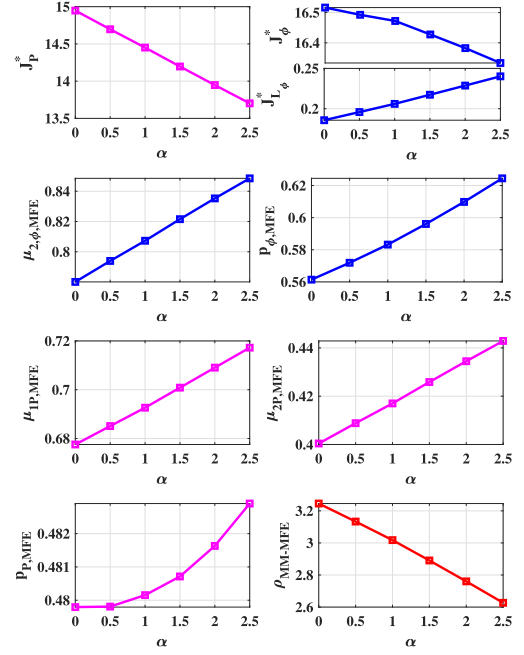


Fig. 15. Variation of the equilibrium policies, the equilibrium costs and the equilibrium load for the primary and the secondary users with the price parameter α .

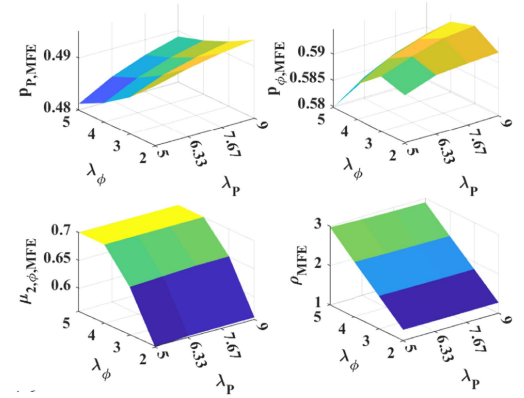


Fig. 16. Variation of the equilibrium policies of the primary and secondary users arrival rates λ_P and λ_ϕ .

the change in the equilibrium load ρ_{MFE} which affects the final equilibrium values. The value of ρ_{MFE} shows a decrease due to the lesser and lesser affinity of the secondary users toward offloading, thereby causing a decrease in the equilibrium load at the primary user's transmitter.

2) *Effect of λ_P and λ_ϕ on MFE:* Next, in Fig. 16, we study the effect of variations in primary and secondary users' incoming arrival rates λ_P , λ_ϕ , respectively, on the MFE policies. We take $V = 10$, $\eta = 0.5$, $\alpha = 1$, $\mu_3 = 15$, $P_{max} = 2$, $f_{\phi,max} = 0.7$, $f_{P,max} = 0.5$, and $N = 30$. Furthermore, the initial conditions were set to $(\mu_{1P}, \mu_{2P}, p_P) = (0.6, 0.3, 0.5)$ and $(\mu_{2,\phi}, p_\phi) = (0.2, 0.5)$. First, we observe from the bottom left plot that as the arrival rates increase, the secondary user's local processor operating frequency at equilibrium, $\mu_{2,\phi,MFE}$, increases to accommodate the increasing incoming tasks. Further, the probability $p_{\phi,MFE}$ of serving locally shows a nominal decrease (from the top

right plot) to compensate for increasing preemptions at the local processor due to increased arrival rate λ_ϕ . The primary user's probability of serving locally also shows a nominal increase to accommodate the increasing arrivals caused by the increasing average AoI cost as a result of increased preemptions at the transmitter of the primary user. Finally, the mean-loading at equilibrium, ρ_{MFE} , increases with increasing arrival rate of the secondary users while being almost unaffected by the increasing arrival rate of the primary user (as the mean effect is only generated by the population of secondary users).

3) *Effective resource utilization*: In our final experiment, we study the effective resource utilization with and without the presence of secondary users. We take $V = 10$, $\eta = 0.5$, $N = 30$, $\mu_3 = 15$, $f_{P,\max} = 0.5$, $P_{\max} = 2$, and $\lambda_P = 2$. First, we note that for the case where only the primary user used its transmitter T_P , and hence, the ES, the effective busy period of T_P was obtained as $t_{T_P,1} = 0.6248$ by solving for the device's optimization problem (as defined in Problem 3) with $\lambda_s = 0$ and $\alpha = 0$. Thus, T_P was idling $\sim 38\%$ of the time. However, for the case, when there are secondary users (with $\alpha = 1$, $f_{\phi,\max} = 0.7$, and $\lambda_\phi = 1$) which can utilize the ES through T_P , the effective busy period of T_P was calculated as $t_{T_P} = t_{T_P,1} + t_{T_P,2} = 0.6032 + 0.3864 = 0.9896$, where $t_{T_P,1}$, $t_{T_P,2}$ were defined above Problem 3. Thus, the idling time is only $\sim 1\%$ in this case. The effective utilization of the shared transmitter has thus increased by a significant 37%, leading to a better utilization of the ES, which is in line with our motivation behind the primary-secondary based-MEC system with priority access to begin with, which was inspired by the CRN technology.

IX. DISCUSSION & CONCLUSION

In this work, we have considered two different MEC architectures with and without priority access, comprised of N devices where each device is assisted by an edge server to service computation intensive tasks. To alleviate the issues of scalability under a large population of users, we have provided low complexity algorithms for both the architectures to compute fully distributed approximate Nash equilibrium policies for each device by setting up the optimization problem of each device toward balancing the power consumed as a result of local processor usage and task offloading, and the timeliness of information affected by the usage of a shared edge computing facility in the presence of a large number of other users. In the process, we invoked techniques from stochastic hybrid systems theory to obtain necessary equations to characterize the average AoI of each user, and consequently employed the mean-field game (and major-minor mean-field game) paradigms to allow for distributed decision making at each user by making use of local information and the statistical information about the system. Finally, we performed extensive numerical evaluations, leading to insights into how system parameter variations affect the equilibrium offloading policies for the users, the average AoI incurred, and the power consumed by the devices.

Our results pave a way for exciting future research directions. The first one that deserves mention is to relax the discipline of LCFS-P to ones with the presence of (a possibly finite capacity)

buffer facility for the users. This would then allow for a packet in service to be processed and delivered while other packets still wait in the queue. Considering such a model, however, makes the AoI analysis more complex due to the increase in state space of the overall MEC system with the length of the buffer. It would be interesting to examine how this impacts the users' equilibrium offloading policies as we explore new trade-offs introduced by the buffer length. Furthermore, it would also be interesting to consider non-stationary user type distribution such as those based on Markovian transitions, redundant sensing, and robustness of the offloading policies against dynamic network fluctuations and misbehaving users, both of which can be taken into account using trust-based models and incentive compatible mechanism design techniques.

Another important direction would be to consider a networked MEC architecture where each user can reach the ES through neighbouring devices' transmitters over multiple hops. This would then lead to interesting notions of device-to-device/peer-to-peer communication within a joint competitive-cooperative game framework, whereby each device wants to balance its selfish objectives whilst also requiring to cooperate with other users by allowing them to utilize their transmitters for establishing communication with the ES. Such a problem formulation may help address the challenges of how to price external device utilization, how to discourage free riding, and how to obtain Nash equilibrium policies for each user within a large user setting. One of the promising frameworks for this setting is that of graphon mean-field games [51], which are counterparts of *standard* mean-field games employed in this work. The former allows for heterogeneous interactions between agents occurring over a network which can be modeled by using techniques from graph theory and consequently attempts to compute equilibrium decision policies for each user in the game.

REFERENCES

- [1] S. Aggarwal, M. A. uz Zaman, M. Bastopcu, S. Ulukus, and T. Başar, "Fully decentralized task offloading in multi-access edge computing systems," in *Proc. IEEE Globecom Workshops*, 2024, pp. 1–6.
- [2] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1628–1656, Thirdquarter 2017.
- [3] M. Hua, Y. Huang, Y. Wang, Q. Wu, H. Dai, and L. Yang, "Energy optimization for cellular-connected multi-UAV mobile edge computing systems with multi-access schemes," *J. Commun. Inf. Netw.*, vol. 3, no. 4, pp. 33–44, 2018.
- [4] A. Muhammad, I. Sorkhoh, M. Samir, D. Ebrahimi, and C. Assi, "Minimizing age of information in multiaccess-edge-computing-assisted IoT networks," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13052–13066, Aug. 2022.
- [5] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update," in *Proc. IEEE Int. Conf. Comput. Commun.*, Mar. 2012, pp. 2731–2735.
- [6] M. Huang, P. E. Caines, and R. P. Malhamé, "Uplink power adjustment in wireless communication systems: A stochastic control analysis," *IEEE Trans. Autom. Control*, vol. 49, no. 10, pp. 1693–1708, Oct. 2004.
- [7] M. Huang, R. P. Malhamé, and P. E. Caines, "Large population stochastic dynamic games: Closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle," *Commun. Inf. Syst.*, vol. 6, no. 3, pp. 221–252, 2006.
- [8] J. M. Lasry and P. L. Lions, "Mean field games," *Japanese J. Math.*, vol. 2, no. 1, pp. 229–260, 2007.

- [9] Y. Wang, F. R. Yu, H. Tang, and M. Huang, "A mean field game theoretic approach for security enhancements in mobile ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1616–1627, Mar. 2014.
- [10] S. Aggarwal, M. A. U. Zaman, M. Bastopcu, and T. Başar, "Weighted age of information based scheduling for large population games on networks," *IEEE J. Sel. Areas Inf. Theory*, vol. 4, pp. 682–697, 2023.
- [11] R. Priyadarshi, R. R. Kumar, and Z. Ying, "Techniques employed in distributed cognitive radio networks: A survey on routing intelligence," in *Proc. Multimedia Tools Appl.*, 2024, pp. 1–52.
- [12] D. Firoozi and P. E. Caines, "The execution problem in finance with major and minor traders: A mean field game formulation," in *Proc. Adv. Dynamic Mean Field Games, Theory, Appl., Numer. Methods*, 2017, pp. 107–130.
- [13] Y. Mao, J. Zhang, S. Song, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," in *Proc. IEEE Glob. Commun. Conf.*, 2016, pp. 1–6.
- [14] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1784–1797, Mar. 2018.
- [15] S. Mao, S. Leng, S. Maharjan, and Y. Zhang, "Energy efficiency and delay tradeoff for wireless powered mobile-edge computing systems with multi-access schemes," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1855–1867, Mar. 2020.
- [16] Y. Jia, C. Zhang, Y. Huang, and W. Zhang, "Lyapunov optimization based mobile edge computing for Internet of Vehicles systems," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7418–7433, Nov. 2022.
- [17] M. Neely, *Stochastic Network Optimization With Application to Communication and Queueing Systems*. Berlin, Germany: Springer, 2022.
- [18] Q. Kuang, J. Gong, X. Chen, and X. Ma, "Age-of-information for computation-intensive messages in mobile edge computing," in *Proc. IEEE Conf. Wireless Commun. Signal Process.*, Oct. 2019, pp. 1–6.
- [19] L. Liu, X. Qin, X. Xu, H. Li, F. R. Yu, and P. Zhang, "Optimizing information freshness in MEC-assisted status update systems with heterogeneous energy harvesting devices," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 17057–17070, Dec. 2021.
- [20] L. Lin, X. Liao, H. Jin, and P. Li, "Computation offloading toward edge computing," *Proc. IEEE*, vol. 107, no. 8, pp. 1584–1607, Aug. 2019.
- [21] C. Feng, P. Han, X. Zhang, B. Yang, Y. Liu, and L. Guo, "Computation offloading in mobile edge computing networks: A survey," *J. Netw. Comput. Appl.*, vol. 202, 2022, Art. no. 103366.
- [22] S. Zhou and W. Jadoon, "The partial computation offloading strategy based on game theory for multi-user in mobile edge computing environment," *Comput. Netw.*, vol. 178, 2020, Art. no. 107334.
- [23] Z. Ning et al., "Mobile edge computing enabled 5G health monitoring for Internet of Medical Things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 463–478, Feb. 2021.
- [24] X.-Q. Pham, T. Huynh-The, E.-N. Huh, and D.-S. Kim, "Partial computation offloading in parked vehicle-assisted multi-access edge computing: A game-theoretic approach," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10220–10225, Sep. 2022.
- [25] H. Teng, Z. Li, K. Cao, S. Long, S. Guo, and A. Liu, "Game theoretical task offloading for profit maximization in mobile edge computing," *IEEE Trans. Mobile Comput.*, vol. 22, no. 9, pp. 5313–5329, Sep. 2023.
- [26] H. Zhou, Z. Wang, N. Cheng, D. Zeng, and P. Fan, "Stackelberg-game-based computation offloading method in cloud-edge computing networks," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16510–16520, Sep. 2022.
- [27] M. Huang, P. E. Caines, and R. P. Malhamé, "Large-population cost-coupled LQG problems with nonuniform agents: Individual-mass behavior and decentralized ϵ -Nash equilibria," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1560–1571, Sep. 2007.
- [28] Y. Kang, Y. Zhu, D. Wang, Z. Han, and T. Başar, "Joint server selection and handover design for satellite-based federated learning using mean-field evolutionary approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 2, pp. 1655–1667, Mar./Apr. 2024.
- [29] S. Y. Olmez et al., "Modeling presymptomatic spread in epidemics via mean-field games," in *Proc. Amer. Control Conf.*, 2022, pp. 3648–3655.
- [30] R. Carmona, "Applications of mean field games in financial engineering and economic theory," 2020, *arXiv:2012.05237*.
- [31] A. Y. Al-Zahrani, F. R. Yu, and M. Huang, "A joint cross-layer and colayer interference management scheme in hyperdense heterogeneous networks using mean-field game theory," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1522–1535, Mar. 2016.
- [32] N. Irtija, I. Anagnostopoulos, G. Zervakis, E. E. Tsiropoulou, H. Amrouch, and J. Henkel, "Energy efficient edge computing enabled by satisfaction games and approximate computing," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 1, pp. 281–294, Mar. 2022.
- [33] R. A. Banez et al., "Mean-field-type game-based computation offloading in multi-access edge computing networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8366–8381, Dec. 2020.
- [34] R. D. Yates, "Status updates through networks of parallel servers," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018, pp. 2281–2285.
- [35] S. K. Kaul and R. D. Yates, "Timely updates by multiple sources: The M/M/1 queue revisited," in *Proc. 54th Annu. Conf. Inf. Sci. Syst.*, 2020, pp. 1–6.
- [36] N. Sathyavageswaran, R. D. Yates, A. D. Sarwate, and N. Mandayam, "Timely offloading in mobile edge cloud systems," in *Proc. IEEE Inf. Theory Workshop*, 2024, pp. 133–138.
- [37] F. Chiariotti, "Age of information analysis for a shared edge computing server," *IEEE Trans. Commun.*, vol. 72, no. 12, pp. 7826–7841, Dec. 2024.
- [38] Y. Gai, H. Liu, and B. Krishnamachari, "A packet dropping mechanism for efficient operation of M/M/1 queues with selfish users," *Comput. Netw.*, vol. 98, pp. 1–13, 2016.
- [39] M. Bastopcu and S. Ulukus, "Age of information for updates with distortion: Constant and age-dependent distortion constraints," *IEEE/ACM Trans. Netw.*, vol. 29, no. 6, pp. 2425–2438, Dec. 2021.
- [40] B. Buyukates and S. Ulukus, "Timely distributed computation with stragglers," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5273–5282, Sep. 2020.
- [41] H. Li, J. Zhang, H. Zhao, Y. Ni, J. Xiong, and J. Wei, "Joint optimization on trajectory, computation and communication resources in information freshness sensitive MEC system," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 4162–4177, Mar. 2024.
- [42] X. Song, X. Qin, Y. Tao, B. Liu, and P. Zhang, "Age based task scheduling and computation offloading in mobile-edge computing systems," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshop*, 2019, pp. 1–6.
- [43] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.
- [44] Y. Sun, I. Kadota, R. Talak, and E. Modiano, *Age of Information: A New Metric for Information Freshness*. Berlin, Germany: Springer, 2022.
- [45] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [46] J. P. Hespanha, "Modelling and analysis of stochastic hybrid systems," *IEEE Proc. Control Theory Appl.*, vol. 153, no. 5, pp. 520–535, 2006.
- [47] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, Mar. 2019.
- [48] J. R. Norris, *Markov Chains*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [49] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA, USA: SIAM, 1998.
- [50] S. J. Wright, "Coordinate descent algorithms," *Math. Program.*, vol. 151, no. 1, pp. 3–34, 2015.
- [51] P. E. Caines and M. Huang, "Graphon mean field games and their equations," *SIAM J. Control Optim.*, vol. 59, no. 6, pp. 4373–4399, 2021.



Shubham Aggarwal (Graduate Student Member, IEEE) received the BTech and MTech degrees in electrical engineering from the Indian Institute of Technology (BHU), Varanasi, India, and the MS in mathematics from the University of Illinois Urbana Champaign, IL, USA, where he is currently working toward the PhD degree with Coordinated Science Laboratory and also with MechSE Department. His current research interests include semantic communication methods and large multi-agent systems.



Muhammad Aneeq uz Zaman received the BS degree in mechatronics engineering from the National University of Science and Technology (NUST), Pakistan, and the MSc and PhD degrees in mechanical engineering from the University of Illinois Urbana-Champaign. He is currently a senior research scientist with Analog Devices Inc. His research focuses on multi-agent reinforcement learning (MARL) using the mean-field game paradigm among others.



Melih Bastopcu (Member, IEEE) received the BS degree in electrical and electronics engineering from Bilkent University, Türkiye, in 2016, and the MSc and PhD degrees in electrical engineering from the University of Maryland at College Park, MD, USA, in 2020 and 2021, respectively. He was a postdoctoral research associate with Coordinated Science Laboratory, University of Illinois Urbana-Champaign, between 2021 and 2024. He is currently an assistant professor with the Department of Electrical and Electronics Engineering, Bilkent University, Türkiye. His research interests include age of information, resource allocation, low latency systems, communication network design, dissemination of misinformation, and social networks. In 2024, he was awarded the prestigious TÜBİTAK 2232-B Fellowship. Appointed in 2025, he is an associate editor for *IEEE Open Journal of the Communications Society*.



Sennur Ulukus (Fellow, IEEE) received the BS and MS degrees in electrical and electronics engineering from Bilkent University, and the PhD degree in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University. She was the senior technical staff member with AT&T Labs-Research. She is a currently a distinguished university professor and Anthony Ephremides professor of information sciences and systems with the Department of Electrical and Computer Engineering, University of Maryland at

College Park, where she also holds a joint appointment with the Institute for Systems Research (ISR). Dr. Ulukus is also a distinguished scholar-teacher with the University of Maryland. Her research interests include information theory, wireless communications, machine learning, signal processing and networks, with recent focus on private information retrieval, age of information, machine learning for wireless, distributed coded computing, group testing, physical layer security, energy harvesting communications, and wireless energy and information transfer. She was the recipient of 2003 IEEE Marconi Prize Paper Award in Wireless Communications, 2019 IEEE Communications Society Best Tutorial Paper Award, 2020 IEEE Communications Society Women in Communications Engineering (WICE) Outstanding Achievement Award, 2020 IEEE Communications Society Technical Committee on Green Communications and Computing (TCGCC) Distinguished Technical Achievement Recognition Award, 2005 NSF CAREER Award, 2011 ISR Outstanding Systems Engineering Faculty Award, and 2012 ECE George Corcoran Outstanding Teaching Award. She was a distinguished lecturer of IEEE Information Theory Society from 2018 to 2019.



Tamer Başar (Life Fellow, IEEE) received the BSEE degree from Robert College, Istanbul, and the MS, MPhil, and PhD degrees from Yale University. He was the interim director of Beckman Institute from 2008 to 2010, interim dean of engineering, in 2018, and director of the Center for Advanced Study (CAS), University of Illinois Urbana-Champaign, from 2014 to 2020. Since 1981, he has been with the University of Illinois Urbana-Champaign, where he is currently the Swanlund Endowed Chair Emeritus and Center for Advanced Study (CAS) professor Emeritus of electrical and computer engineering, with also affiliations with the Coordinated Science Laboratory, Information Trust Institute, and Mechanical Science and Engineering. He has more than 1000 publications in systems, control, communications, optimization, networks, and dynamic games, including books on non-cooperative dynamic game theory, robust control, network security, wireless and communication networks, and stochastic networked control. He is the editor of several book series. His current research interests include stochastic teams, games, and networks, risk-sensitive estimation and control, mean-field game theory, multi-agent systems and learning, data-driven distributed optimization, epidemics modeling and control over networks, strategic information transmission, spread of disinformation, and deception, security and trust; energy systems, and cyber-physical systems. He is a member of the US National Academy of Engineering and fellow of the American Academy of Arts and Sciences, IFAC, and SIAM. He was the president of IEEE CSS (Control Systems Society), ISDG (International Society of Dynamic Games), and AACC (American Automatic Control Council). He was the recipient of several awards and recognitions more than the years, including the highest awards of IEEE CSS, IFAC, AACC, and ISDG, IEEE Control Systems Award, 2021 the Wilbur Cross Medal from Yale University, and a number of international honorary doctorates and professorships. He was the editor-in-chief of *Automatica* between 2004 and 2014.