IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 17, 2020, accepted March 11, 2020, date of publication March 18, 2020, date of current version March 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981837

Cognitive Data Offloading in Mobile Edge Computing for Internet of Things

PAVLOS ATHANASIOS APOSTOLOPOULOS[1], (Student Member, IEEE),
EIRINI ELENI TSIROPOULOU[1], (Member, IEEE), AND
SYMEON PAPAVASSILIOU[2], (Senior Member, IEEE)
[1]Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131, USA
[2]School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Zografou, Greece

Corresponding author: Eirini Eleni Tsiropoulou (eirini@unm.edu)

This work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) through the First Call for H.F.R.I. Research Projects to support Faculty Members and Researchers and the Procurement of High-Cost Research Equipment Grant under Project HFRI-FM17-2436. The work of Eirini Eleni Tsiropoulou was supported in part by the NSF under Grant CRII-1849739.

ABSTRACT Data offloading to Mobile Edge Computing (MEC) servers is an attractive choice for resource-constrained Internet of Things (IoT) devices, towards reducing their computational effort. In this paper, we investigate the potential of partial data offloading to MEC servers, under the perspective of users' cognitive IoT devices presenting loss averse and gain seeking behavior. Due to the sharing nature of the access environment and the MEC server's computational characteristics, we treat the MEC server option as a common pool of resources with uncertain payoff returned to the users, while the local computation capability is treated as a safe option for each user. Following the properties of Prospect Theory, users' prospect-theoretic utilities are formulated exploiting the local computing and offloading overhead options under probabilistic uncertainty. Such a modeling allows for the infusion of human awareness, inherent cognitive biases and behavioral characteristics into the devices' operation, their data offloading decisions and the edge computing environment that the devices are interacting with. Accordingly, each user's optimal offloaded data to the MEC server is obtained as the outcome of a non-cooperative game, with users attempting to maximize their own utilities. The existence and uniqueness of a Pure Nash Equilibrium (PNE) are proven under the probabilistic nature of the respective payoff functions, while a distributed algorithm that convergences to the PNE is designed. Numerical results are provided that demonstrate the operation and superiority of the proposed framework under different IoT scenarios and behaviors, considering both homogeneous and heterogeneous users.

INDEX TERMS Intelligent data offloading, mobile edge computing, Internet of Things, risk-based behavior modeling, cognitive decision making, probabilistic uncertainty.

I. INTRODUCTION

The rise of Internet of Things (IoT) has provided great benefits to people by creating a system of interrelated computing, sensing, and communication devices that facilitates and improves humans' every-day life. IoT is foreseen to reach 500 billion devices that are connected to the Internet by 2030 [1], while the global mobile traffic is expected to increase sevenfold by 2021 [2]. Till recently, to satisfy the computation and storage requirements of IoT, cloud computing has served as the most important computing infrastructure. However, with the explosion of the number of devices in IoT and the large volume of the generated data, cloud computing has been proven inadequate to efficiently handle the corresponding loads, while meeting the user requirements in terms of delay/latency and energy efficiency. Therefore, edge computing - representing the practice of processing data near the edge of the network [3], [4]- is gaining significant momentum as complementary computing paradigm that overcomes the aforementioned challenges.

In particular, Mobile Edge Computing (MEC) is becoming a key flexible and cost-effective mechanism that enables the mobile devices to offload their computation tasks to servers residing at the ''edge'' of the radio access networks [5], [6]. MEC reduces the computational effort of the devices, which are usually characterized by limited memory, computational

The associate editor coordinating the review of this manuscript and approving it for publication was Min Jia.

capability, and battery life. Given that the MEC allows the devices to offload their computation tasks to a MEC server typically through the wireless access network, both the computation and communication challenges should be jointly studied [7]. It is noted that in the recent literature [5] the term data offloading has been used interchangeably with computation offloading. Indeed, a computation task is considered to consist of a set of data that the user needs to process and part (or all) of them are offloaded to the MEC server. In the rest of this paper, following this trend, and for simplicity, mainly the term of data offloading is adopted.

In parallel, a key observation is that computing systems have evolved over the years from imperative computing, to autonomic computing, and to cognitive computing [8]. Initially autonomic computing aims at automatically adapting the system behaviors based on its context changes, while cognitive computing introduces intelligent systems capable of perceiving, learning and thinking as close as possible to human patterns. This vision along with the proliferation of Internet of Things (IoT) has driven and motivated the Cognitive IoT. Our current work is well aligned with this development and evolution, by properly considering the infusion of human awareness and behavioral characteristics into the devices, their data offloading decisions, and the edge computing environment that the devices are interacting with.

Furthermore, data offloading can be in general classified into three categories, namely: (a) always offload; (b) all or nothing offloading, where either the entire data is offloaded or the entire data is processed locally, with the offloading decision typically to depend on energy thresholds; and (c) partial offloading, where some parts are offloaded with the remaining to be executed locally. Our research work focuses on the latter category, as it offers the greatest flexibility and potential for intelligence and optimization, based on both communication and computation environment awareness.

Specifically, in this paper, given that edge computing tends to introduce certain communication and computation probabilistic uncertainties due to its shared nature, the focus is placed on the problem of efficient resource management and intelligent (partial) data offloading approaches in edge computing for the Internet of Things. A key novel characteristic of our work, is that the overall resource management and offloading decision making process, is performed under the adoption and consideration of a realistic cognitive behavioral paradigm for the involved IoT entities. The latter is driven by and reflects the risk-based behavioral patterns and reactions of the humans they aim to serve, a recently emerged trend and ambition [9].

A. RELATED WORK AND MOTIVATION

Centralized and *distributed* approaches have been proposed in the recent literature to jointly consider the computation and communication limitations in the MEC environment within IoT era [10]. Mao *et al.* [11] proposed a *centralized* joint radio and computational resource management scheme for multi-user MEC systems to minimize the long-term average weighted total devices' and MEC server's power consumption. Also, they have examined a stochastic model of devices' computation task requests, targeting at minimizing the overall power consumption in the system and examining the tradeoff between the devices' power consumption and the computing tasks' execution delay [12]. Muñoz *et al.* [13] proposed to minimize the affordable latency in executing an application in a femto-cloud computing environment by exploiting the trade-off between energy consumption and latency. In [14] and [15], the authors consider orthogonal frequency division multiple access (communication aspect) and devices' workload offloading priorities (computing aspect) aiming at minimizing the weighted sum devices' energy consumption, under the constraint of computation latency. In [16], [17], the authors aim at minimizing only the system's energy consumption in a single MEC server environment, thus concluding to energy-efficient data offloading via jointly examining the computation offloading and the radio resource allocation for all the devices in an IoT-based network. Want *et al.* [18] study the joint allocation of computation and communication resources under two different perspectives, i.e., jointly minimizing the devices' energy consumption and the latency of application execution. Guo *et al.* [19] introduce an energy efficient dynamic offloading and resource scheduling policy to decrease the devices' energy consumption and shorten the computation task completion time. A holistic approach is considered in [20] to minimize the overall cost of energy, computation and delay for the devices, while jointly optimizing their offloading decisions and the allocation of the computation and communication resources.

On the other hand, *distributed* resource management approaches have been proposed to support the devices' autonomous behaviour and reduce the devices' and MEC server's signaling overhead. The recently emerged worlds of networked human-driven devices, e.g., smartphones, and networked things, e.g., networked appliances and sensors, have already started to be treated in a unified manner. The mobile devices and relevant software are at a large extent designed to reflect the wishes and decision-making of the users they serve. Thus, the human factor (i.e., intelligence, personality, behavior and social structures) has strongly shaped the requirements and capabilities of the IoT devices, while infusing to them a certain level of ability to recognize, perceive and exhibit behavioral patterns [9]. Therefore, the autonomous and distributed decision-making of the IoT devices in the MEC IoT environment is of utmost importance.

Specifically, Chen [21] studied the decentralized computation offloading decision-making problem among the devices in a MEC environment and formulated a distributed computation offloading game among them to decide if the computing task will be performed locally or at the MEC server. This work has been extended in [7] for a multi-channel wireless communication environment. In [22], the decision-making problem of either offloading the devices' computation tasks to multiple MEC servers or performing the computation locally has been formulated as a non-cooperative game among the devices and a Nash equilibrium is determined. In [23] and [24], novel incentive-based approaches have been proposed to improve the offloading efficiency via introducing pricing policies to encourage the fair and high quality MEC services' sharing by the devices. The problem of determining the number of MEC servers to maximize the devices' utilities, which are expressed through their Quality of Service (QoS) prerequisites is studied in [25] adopting the minority games.

Moreover, the authors in [26] define a computation task offloading problem, aiming to determine which components of the user's computation task should be offloaded to the mobile cloud computing provider in order to minimize the execution cost. The latter is assumed to consist of the local computing cost on the user's device, the uploading cost from the user to the cloud, and the downloading cost from the cloud to the user's device. In [27], a constrained optimization problem is formulated to determine the users' optimal data offloading while considering the latency and reliability constraints in an ultra-reliable and low latency communication network, and is solved based on the Lyapunov stochastic optimization approach. In [28], the authors introduce a mathematical model to capture the computation offloading cost, i.e., the time and energy consumption, in a mobile cloud computing environment, which is further reduced in [29] by using a framework for instruction translation and offloading, while considering multimedia-based applications.

The aforementioned approaches, either centralized or distributed, assume that the users and their corresponding IoT devices have a risk-neutral behavior, acting as neutral maximizers that aim to maximize their payoff from the joint allocation of the communication and computation resources. However, in real life, the individuals and the IoT devices that present cognitive behavior mimicking their owners' behavioral patterns, tend to exhibit risk-seeking or loss-aversion behavior under uncertainty, which is a key property of the MEC environment. Therefore, a first key step towards properly realizing the vision of Cognitive IoT encompasses the infusion of more human awareness into the devices and environments we interact with. While computers are not yet capable of general human-like thought, they can now perform some of the same underlying functions that humans perceive in their decision making process. Towards identifying and rigorously studying cognitive biases with this emerging era, it has been argued [30] that a simple version of expected utility theory does not properly describe human behavior. Instead, Prospect Theory [31] has emerged as a realistic model of how people make decisions, by successfully modeling and considering many of their standard biases [32].

B. CONTRIBUTIONS AND OUTLINE

In order to fill the aforementioned research gap, in this paper we exploit Prospect Theory [31] to capture users' realistic choices in terms of loss aversion and gain seeking characteristics within the MEC environment, when they make decisions about data offloading. Such a consideration allows us to overcome the drawbacks associated with the majority of the relevant research efforts that have considered risk-neutral users and classical utility maximization approaches.

The use of Prospect Theory in our work is well motivated and supported by the observations that MEC IoT environment presents a competitive resource-constrained environment, where the users are making decisions under uncertainty, which stems from the partial available information and the competition to share the limited resources. The individual entities of a system, i.e., cognitive IoT devices, make distributed and autonomous decisions under risk and uncertainty of the associated payoff of their decisions, which is determined in a probabilistic manner, while they may demonstrate systematic deviations from the expected utility theory, where all the individuals are assumed as risk neutral with respect to their choices. Prospect Theory has already been applied in various applications, including cyber-physical social systems [33], dynamic resource management in 5G wireless networks [34], [35], UAV-assisted communications in public safety networks [36], [37], and anti-jamming communications in cognitive radio networks [38]. To the best of our knowledge, the proposed framework constitutes the first effort towards modeling and realizing the user risk-based data offloading behavior and decision making, in a MEC IoT environment under uncertainty, thus offering a holistic, cognitive and risk-aware approach.

The main contributions of our work that differentiate it from the rest of the literature are summarized below:

1. The total overhead, in terms of time delay and energy consumption regarding the local computing and offloading choices, is introduced (Section II-A) and exploited to define the devices' actual utility functions (Section II-B).
2. The cognitive-enabled devices decide under risk the computation load to be offloaded to the MEC server, given the computation and communication uncertainty due to the MEC server environment shared nature, in terms of computation competition and access interference. In that respect, the devices' decision-making is influenced and shaped by the key prospect-theoretic characteristics (Section III-A). The choice of locally performing the computation tasks is considered as a ''safe'' option offering predictable performance and satisfaction, while the MEC server is treated as a common pool of resources (CPR), providing an uncertain payoff to the devices (Section III-B).
3. Devices' prospect-theoretic utility functions are properly formulated by considering their actual utilities, their computation tasks, atheir cognition biases and their reflection to gains and losses in the MEC environment (Section III-C). The problem of each device determining in an autonomous manner the portion of computation task that will be performed at the MEC server (CPR), has been formulated as an optimization problem of each device's expected prospect-theoretic utility, and treated as a non-cooperative game among the devices (Section IV).
4. The non-cooperative game is solved in a distributed manner and the existence and uniqueness of a pure Nash

equilibrium (PNE) is shown (Section IV-B). It is noted that this goal becomes challenging due to the probabilistic nature of the payoff function, which differentiates the solution of our problem compared to the vast majority of the literature on non-cooperative games. A distributed and low-complexity algorithm that converges to the PNE is also introduced (Section IV-C).
5. A series of experiments are performed to evaluate the performance and the inherent attributes of the proposed device-centric risk-based data offloading decision-making framework (Section V). A comparative study demonstrates its superiority and benefits, in terms of user satisfaction and proper system operation. Finally, Section VI concludes the paper, and highlights some interesting open issues of high research and practical importance.

II. SYSTEM MODEL

A. COMPUTATION AND COMMUNICATION MODEL

A set of $\aleph = \{1, \ldots, i, \ldots, N\}$ collocated devices is considered, where each device $i \in \aleph$ has a computational intensive task T_i to be completed. Furthermore, we consider the uplink of a wireless network, consisting of a base station (BS) acting as a MEC server, with an upper bounded computation capability for task execution. We consider a quasi-static scenario, where the set of devices remains unchanged during a computation offloading period.

Data partitioned oriented applications are considered where each device $i \in \aleph$ has a computation task $T_i = (I_i, C_i)$, where I_i and C_i denote the computation input bits and the total number of CPU cycles required to accomplish the computation task T_i, respectively. We consider $C_i = \lambda_i * I_i$, where the parameter λ_i ($\lambda_i > 0$) expresses the computational complexity of the task requested by the device i, $i \in \aleph$ and its value depends on the nature of the application, e.g., a higher λ_i expresses a more computation intensive task. We assume that each computation task T_i, can be arbitrarily partitioned into subsets of any size, so each device can offload an amount of data $b_i \in [0, I_i]$ to the MEC server and keep the rest for local computing. We have $b_i = 0$, if user $i \in \aleph$ decides to compute its whole task locally. We consider a typical interference limited communication environment, where the MEC server is the receiver of the users' transmitted data and each user experiences the interference imposed by the transmissions of the rest of the users in the examined MEC IoT environment. Given the decision profile $\mathbf{b} = [b_1, b_2, \ldots, b_N]$ of all users, the uplink data rate for the computation offloading of device i is [39]:

$$R_i = W * log_2(1 + \frac{P_i * G_i}{\sigma^2 + \sum_{j=1, b_j \neq 0, j \neq i}^{N} P_j * G_j}) \tag{1}$$

where W denotes the system's bandwidth, P_i the user's i transmission power, G_i the channel gain between the device i and the BS, and σ^2 is the background noise. An overview of the overall prospect-theoretic data offloading in a mobile edge computing cognitive-enabled IoT environment is depicted in Fig. 1. Also, a summary of the key notation adopted in this paper is presented in Table 1.

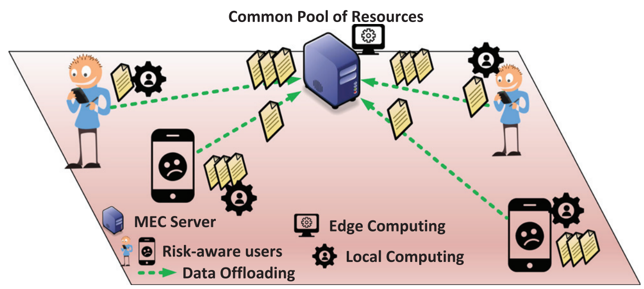


FIGURE 1. Prospect-theoretic data offloading in mobile edge computing.

TABLE 1. Summary of key notations.

Notation	Description
$\aleph$	Set of users IoT devices
T_i	Computation task
I_i	Computation input bits
C_i	Total number of required CPU cycles
λ_i	Task's computational complexity
b_i	Amount of offloaded data
R_i	Uplink data rate
W	System's bandwidth
P_i	Transmission power
G_i	Channel gain
σ^2	Background noise
$\mathbf{b}$	Data offloading vector for all users
$O_i^{f,e}$	Energy consumption overhead
$O_i^{f,tr}$	Transmission time overhead
F_i^f	MEC server's computing resources devoted to user i
$O_i^{f,t}$	Execution time of the offloaded data
O_i^f	Total offloading overhead for user i
w_i^t, w_i^e	Weights of time delay and energy consumption overheads
f_i	Consumed energy per CPU cycle
$O_i^{l,t}$	Local execution time overhead
F_i^l	Local computation capability
$O_i^{l,e}$	Local energy consumption overhead
O_i^l	Total local computing overhead
O_i	Total experienced overhead
b_T	Total amount of offloaded data by all the users
$d(b_T)$	Production function
F_{MEC}	MEC server's upper bound computation capability
b_{th}	MEC server's received bytes threshold value
z_i	User's actual utility function
$p(b_T)$	Probability of failure
u_i	Prospect-theoretic utility function
$z_{i,0}$	Reference point
α_i, γ_i	User's sensitivity to gains and losses
k_i	Loss aversion parameter
$\mathbb{E}(u_i)$	Expected prospect-theoretic utility function

1) *Offloading Overhead*: A user i offloads $b_i \in [0, I_i]$ amount of data to the MEC server, where the latter executes this part of the task T_i on behalf of the user. The user i has a total offloading overhead consisting of the following terms: a) the energy consumption to transmit the data b_i, b) the transmission time and c) the execution time of the computation task at the MEC server. The energy consumption overhead $O_i^{f,e}$ is determined by the consumed energy during the transmission of the data b_i to the MEC server as follows: $O_i^{f,e} = \frac{b_i * P_i}{R_i}$. The transmission time overhead $O_i^{f,tr}$ is given as: $O_i^{f,tr} = \frac{b_i}{R_i}$. Similarly, the execution time of the offloaded

data b_i depends on the computing resources (rate of return) F_i^f that the MEC server devotes to the computation task of user i, as follows: $O_i^{f,t} = \frac{\lambda_i * b_i}{F_i^f}$. More details about F_i^f are provided in Section II-B.

Therefore, the total offloading overhead for user $i, i \in \aleph$ to offload b_i data can be obtained as follows:

$$O_i^f(\mathbf{b}) = w_i^e * O_i^{f,e} + w_i^t * (O_i^{f,tr} + O_i^{f,t}) \tag{2}$$

where $w_i^t, w_i^e \in [0, 1], w_i^t + w_i^e = 1$, denote the weights of the time delay and energy consumption overheads, respectively, that can be tuned by each user according to different priorities and considerations, e.g., low battery consideration ($w_i^e > w_i^t$) or delay sensitive application ($w_i^t > w_i^e$). It is noted that the normalization of the energy consumption and the time delay overhead is appropriately taken into account in the weights w_i^e and w_i^t, so as both contributions to be treated fairly in terms of their order of magnitude and impact.

2) *Local Computing Overhead*: A user i executes $(I_i - b_i)$ amount of data of its computation task T_i locally on its device. In this case, the user i has a total local computing overhead consisting of the following terms: a) the local execution time overhead and b) the local energy consumption overhead. The local execution time overhead is given as: $O_i^{l,t} = \frac{\lambda_i * (I_i - b_i)}{F_i^l}$, where $\lambda_i * (I_i - b_i)$ is the number of cycles required for the local computation, and F_i^l denotes the local computation capability (CPU cycles per second) of user i. Similarly, the local energy consumption overhead of the user i, is given as: $O_i^{l,e} = f_i * \lambda_i * (I_i - b_i)$, where $f_i \in \mathbb{R}^+$ denotes the consumed energy per CPU cycle. Therefore, the total local computing overhead of the user i is given as follows:

$$O_i^l(b_i) = w_i^t * O_i^{l,t} + w_i^e * O_i^{l,e} \tag{3}$$

Taking into account that a user may offload part of its computation task to the MEC server, based on Eq. 2 and Eq. 3 its total experienced overhead is:

$$O_i(\mathbf{b}) = O_i^l(b_i) + O_i^f(\mathbf{b}) \tag{4}$$

It is highlighted that in Eq. 4 we consider the total overhead that a device experiences by executing part of its computation tasks locally and at the MEC server, as if the two parts are not executed in parallel. If the two parts of the computation task were executed in parallel, we could consider the largest term instead, i.e., $max(O_i^{f,tr} + O_i^{f,t}, O_i^{l,t})$, as the time delay overhead. Thus, Eq. 4 could be written as $O_i(\mathbf{b}) = w_i^t * max(O_i^{f,tr} + O_i^{f,t}, O_i^{l,t}) + w_i^e * (O_i^{f,e} + O_i^{l,e})$ and would not affect the structure of the rest analysis in the paper, which would remain valid.

B. DEVICE's ACTUAL UTILITY

In this section, the users' actual utilities expressing their satisfaction from executing part of their computation task at the MEC server (CPR) and the rest locally at the device, are formulated. The exploitation of the MEC server's computation capabilities via offloading part of the user's computation task to the server provides a corresponding satisfaction to the user, which depends on server's workload. This satisfaction is captured by the rate of return function F_i^f, which is personalized based on each device's task's computational complexity λ_i, and decreases as the total computation offloading by all devices increases due to the upper bounded computation capacity of the MEC server. Specifically, the MEC server provides its computation capabilities to the users in a fair and proportional manner. Thus, the devices whose computation tasks are characterized by higher computational complexity, i.e., λ_i, experience an improved rate of return as the MEC server managed to fulfill their demanding computation tasks. The rate of return function for each device $i, i \in \aleph$ is formulated as:

$$F_i^f(b_T) = \frac{\lambda_i}{\sum_{j=1, b_j \neq 0}^{N} \lambda_j} * d(b_T) \tag{5}$$

where $b_T = \sum_{j=1}^{N} b_j$ is the total amount of offloaded data by all the devices to the MEC server and $d(b_T)$ is the production function of the MEC server expressing its computing performance with respect to the total data offloading. The production function is formulated as follows:

$$d(b_T) = \begin{cases} (1 - \frac{b_T}{b_{th}}) * F_{MEC} & \text{if } b_T \leq b_{th} \\ 0 & otherwise \end{cases} \tag{6}$$

where F_{MEC} [CPU Cycles/sec] denotes the MEC server's upper bound computation capability, that is shared among the different offloaded tasks. The parameter b_{th} denotes the MEC server's received bytes threshold value, where if $b_T \geq b_{th}$ the MEC server is considered unable to execute the offloaded tasks into a specific duration of time, thus it "fails". This concept is well-known in the literature as the "Tragedy of the Commons" [40]. As a result, in this case, it is more beneficial for the device $i \in \aleph$ to execute its whole task T_i locally. The consideration of including the MEC server's received bytes threshold value in our analysis captures the operation of a realistic MEC system, where if the MEC server was overwhelmed with data to process, then it would become over-exploited concluding to increased delays. In that case, the computing services offered by the MEC server to the devices become unsatisfactory and of no value to them.

Proposition 1: The production function $d(b_T)$, and each device's $i \in \aleph$ rate of return function $F_i^f(b_T)$, are strictly decreasing with respect to the total offloaded data b_T.

The above proposition holds true in our environment, given that if the MEC server becomes overloaded, the device's choice of offloading part of its task to the MEC server becomes less beneficial as the user suffers the burden of long computation time delays stemming from the over-exploited MEC server. In the following analysis, without loss of generality and for simplicity in the presentation, we consider $w_i^t = w_i^e = 1/2, \forall i \in \aleph$, thus, each device has equal sensitivity to the time delay and the energy consumption overhead. Each device is associated with an actual utility function formulated as a linear combination of the overhead experienced by executing a part of its computation task to the

MEC server and the rest part locally. Thus, the device's actual utility can be formulated via combining Eq. 2 and Eq. 3, as follows:

$$z_i(\mathbf{b}) = b_i * (\frac{\lambda_i}{F_i^f} + \frac{P_i + 1}{R_i}) + \lambda_i * (I_i - b_i) * (\frac{1}{F_i^l} + f_i) \quad (7)$$

III. THE PROSPECT OF DATA OFFLOADING

A. PARTIAL OFFLOADING UNDER PROSPECT THEORY

In real mobile applications, users do not always adopt risk-neutral behavior, instead they tend to demonstrate different actions under losses or gains with respect to their actual utility. Towards capturing the device-centric risk-based decision-making in the MEC environment, the Prospect Theory is adopted [31]. Following this behavioral model, individuals make decisions under risk and uncertainty of the associated payoff of their choices, which is estimated with some probability. Therefore, the users' actual utility as expressed in Eq. 7, is evaluated with respect to a reference point (reference dependence property) [41]. This reference point is considered as the zero point (i.e., ground truth) of the users' actual utility. Given the reference point and users' offloaded data, they determine their prospect-theoretic probabilistic payoff. In our work, we consider as the reference point of each user the corresponding experienced overhead, if the whole task was locally executed.

Users' prospect-theoretic utility function is a concave function with respect to the user's actual perceived utility above the reference point, a convex function below it, and has a greater slope in losses compared to the gains (loss aversion property), as presented in Fig.2. This formulation is well-aligned with the observation that the users weigh more the losses compared to the gains of the same amount of overhead in terms of dissatisfaction and satisfaction, respectively (diminishing sensitivity property).

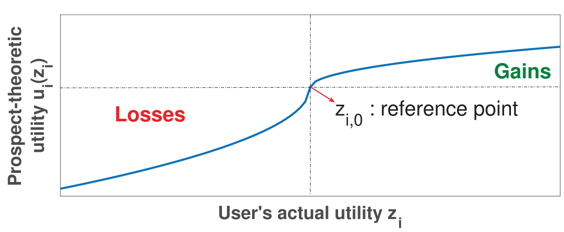


FIGURE 2. Prospect theoretic utility.

B. MEC: A COMMON POOL OF RESOURCES

In the MEC environment under consideration, the MEC server is considered as a Common Pool of Resources (CPR), since it is: a) non-excludable, in the sense that all the users have access to arbitrarily offload their computation tasks to the server, and b) rivalrous and subtractable for the users, as the reservation of computation capabilities by one user from the MEC server, reduces the ability of reserving computation cycles by another [40]. Each user's goal is to determine in an autonomous manner the offloaded data b_i to the CPR with some uncertainty in the expected obtained outcome, while maintaining its remaining amount of data $(I_i - b_i)$ locally to be executed at the device, which is a "safe" computation resource in terms of a priori knowing the total local overhead, so as to minimize its overall perceived overhead. The probability of failure of the MEC server (CPR) is denoted by $p(b_T)$.

Proposition 2: The MEC server (CPR) is characterized by the following properties.

1) The probability of failure $p(b_T)$ is strictly increasing, convex and twice continuously differentiable with respect to $b_T \in [0, b_{th})$, with $p(b_T) = 1, \forall b_T \geq b_{th}$.
2) User's i strategy set of offloading an amount of data to the MEC server is $S_i = [0, \min(I_i, b_{th})], \forall i \in \aleph$.

It is noted that the corresponding probability of failure $p(b_T)$ being strictly increasing with respect to b_T, allows to properly capture the reliability characteristics of the MEC server. Some examples of the MEC server's probability of failure are the logarithmic, the linear, or the exponential function with respect to b_T. The choice of an appropriate function form could be based on various operational factors and characteristics (e.g., the MEC server's robustness to failure, non-linear server's behavior to traffic loads and computing utilization, etc.), and assuming that satisfies the properties in Proposition 2, it does not harm the validity of our analysis. In this paper, without loss of generality and for demonstration purposes only, we consider a linear probability of failure function given as follows: $p(b_T) = \frac{b_T}{b_{th}}$. It is highlighted that in the case of an underloaded MEC server, the probability of failure function will return small values, thus, concluding to better experience and satisfaction for the user that offloads its computation tasks to the MEC server.

C. PROSPECT-THEORETIC UTILITY

Based on the Prospect Theory, the prospect-theoretic utility of a user is defined as follows [42]:

$$u_i(z_i) = \begin{cases} (z_{i,0} - z_i)^{\alpha_i} & \text{if } z_i \leq z_{i,0} \\ -k_i * (z_i - z_{i,0})^{\gamma_i} & \text{if } z_i > z_{i,0} \end{cases} \quad (8)$$

where z_i is the user's $i, i \in \aleph$ actual utility as defined in Eq. 7 and $z_{i,0}$ denotes the reference point of user's prospect-theoretic utility. Each user's reference point $z_{i,0}$ is defined as the actual utility that it experiences by executing its whole task T_i locally at the device.

$$z_{i,0} = z_i|_{b_i=0} = \lambda_i * I_i * (\frac{1}{F_i^l} + f_i) \quad (9)$$

As stated earlier, we have omitted the weight $w_i^t = w_i^e = 1/2$ for simplicity in the presentation. Each device's $i, i \in \aleph$ parameters $\alpha_i, \gamma_i \in (0, 1]$ express the user's sensitivity to the gains and losses of its actual utility z_i, respectively. The risk seeking behavior of a user in losses and its risk averse behavior in gains is reflected by small values of parameter α_i in its prospect-theoretic utility u_i. The small values of parameter γ_i imply higher decrease of user's prospect-theoretic utility for small values of z_i and close to the reference point $z_{i,0}$. In this paper, we consider that the users follow analogous behaviour in losses and gains, thus $\alpha_i = \gamma_i$. Furthermore, the parameter $k_i, k_i \in [0, \infty]$ reflects the impact of losses compared to the gains in users' prospect-theoretic utility. Specifically, if $k_i > 1$, the user i weighs the losses more than the gains, thus it illustrates a loss averse behaviour as it is resistant to lose part of its actual utility z_i. The exact opposite holds true, when $0 \leq k_i \leq 1$, in which case the user weighs the gains more or equal than the losses, thus presenting an aggressive gain seeking behavior. Please note that the use of different parameters α_i, k_i for various users allows taking into account with high granularity all the different characteristics of every user. If a homogeneous population was assumed then we would consider $\alpha_i = \alpha$ and $k_i = k$ for each user $i, i \in \aleph$.

If the MEC server (CPR) does not fail due to the over-offloading of users' data, then each user perceives an actual utility given by the Eq. 7. In this case, the actual perceived utility (overhead) is lower than the reference point $z_{i,0}$, i.e, $z_i \leq z_{i,0}$, as at the reference point the user i executes its whole task T_i locally. Therefore, via subtracting the actual utility z_i (Eq. 7) from the reference point $z_{i,0}$ (Eq. 9) and shaping the result according to the first branch of Eq. 8, we have $u_i(z_i) = [b_i(\frac{\lambda_i}{F_i^l} + \lambda_i f_i - \frac{P_i+1}{R_i} - \frac{\lambda_i}{F_i^f})]^{\alpha_i}$. On the other hand, if the MEC server becomes overloaded and fails to serve the users' offloaded computation tasks, the users' overhead is given by Eq. 7 with $b_i = 0$, as the users have to execute their whole tasks locally, though they experienced the energy consumption and transmission time overhead from offloading b_i data to the MEC server, before the stage of failure is reached. Thus, user's actual utility is $z_i = z_{i,0} + \frac{b_i}{R_i} + \frac{b_i}{R_i} P_i$ and is greater than the reference point $z_{i,0}$. Therefore, by subtracting the reference point from user's actual utility, the second branch of Eq. 8 can be written as $u_i(z_i) = -k_i(b_i \frac{P_i+1}{R_i})^{\alpha_i}$.

Following the aforementioned argumentation we can readily rewrite the user's prospect-theoretic utility (Eq.8) as follows:

$$u_i(z_i) = \begin{cases} [b_i(\frac{\lambda_i}{F_i^l} + \lambda_i f_i - \frac{P_i+1}{R_i} - \frac{\lambda_i}{F_i^f(b_T)})]^{\alpha_i} & \text{if } z_i \leq z_{i,0} \\ -k_i(b_i \frac{P_i+1}{R_i})^{\alpha_i} & \text{if } z_i > z_{i,0} \end{cases} \tag{10}$$

Moreover, for notational convenience we define $\overline{d_i}(b_T) \triangleq (\frac{\lambda_i}{F_i^l} + \lambda_i f_i - \frac{P_i+1}{R_i} - \frac{\lambda_i}{F_i^f(b_T)})^{\alpha_i} > 0$ assuming that the server has not failed and $\epsilon_i \triangleq (\frac{P_i+1}{R_i})^{\alpha_i}$, and therefore Eq. 10 can be written as:

$$u_i(z_i) = \begin{cases} b_i^{\alpha_i} \overline{d_i}(b_T) & \text{if } z_i \leq z_{i,0} \\ -k_i \epsilon_i b_i^{\alpha_i} & \text{if } z_i > z_{i,0} \end{cases} \tag{11}$$

The MEC server's failure to serve the users depends on the total offloaded data by all of them. Given that the probability of the MEC server's failure is $p(b_T)$, the probability that the server survives and executes the offloaded computation tasks is accordingly $(1 - p(b_T))$. As a result, considering the probability of MEC server's failure, Eq. 11 can be written equivalently as follows:

$$u_i(z_i) = \begin{cases} b_i^{\alpha_i} \overline{d_i}(b_T), & \text{with prob. } 1 - p(b_T) \\ -k_i \epsilon_i b_i^{\alpha_i}, & \text{with prob. } p(b_T) \end{cases} \tag{12}$$

Each user's expected prospect-theoretic utility based on all users' offloaded data $\mathbf{b} = [b_1, b_2, \ldots, b_N]$ is given as follows.

$$\begin{aligned} \mathbb{E}(u_i) &= b_i^{\alpha_i} \overline{d}_i(b_T)(1 - p(b_T)) - (k_i \epsilon_i b_i^{\alpha_i}) p(b_T) \\ &= b_i^{\alpha_i} [\overline{d}_i(b_T)(1 - p(b_T)) - k_i \epsilon_i p(b_T)] \\ &\triangleq b_i^{\alpha_i} g_i(b_T) \end{aligned} \tag{13}$$

where $g_i(b_T) = \overline{d}_i(b_T)(1 - p(b_T)) - k_i \epsilon_i p(b_T)$ is the effective rate of return of the MEC server for the user $i, i \in \aleph$.

IV. OPTIMIZING DEVICES' OVERHEAD

A. PROBLEM FORMULATION

The goal of each device is to minimize its perceived overhead from its computation task's execution via sophisticatedly and selfishly offloading part of the task to the MEC server. This problem can be formulated as a maximization problem of each user's prospect-theoretic utility function, as follows:

$$\max_{b_i \in S_i} \mathbb{E}(u_i) = b_i^{\alpha_i} g_i(b_T) \tag{14}$$

In the modeling considered in this work, the latency and energy factors have been directly considered as part of the corresponding overheads, computed for both offloading and local computing cases (i.e., Eq. 2 and Eq. 4), while their combined optimization is treated and achieved through the solution of the optimization problem in Eq. 14. Following the vision of ultra-reliable low latency communications and respecting the energy limitations of the IoT devises, as part of the emerging Tactile Internet [43], this problem could be further extended by considering hard constraints on the required latency of the computation task and the energy availability of the devices. Accordingly, these constraints are expected to reduce the users' strategy space and the corresponding feasible solution space.

The above maximization problem can be confronted as a non-cooperative game among the users who act as players making the optimal decisions about themselves in a selfish and distributed manner. Let $G = [\aleph, \{S_i\}, \{\mathbb{E}(u_i)\}]$ denote the non-cooperative game among the N users, where each user's strategy space is $S_i = [0, min(I_i, b_{th})]$ and its payoff is its expected prospect-theoretic utility function $\mathbb{E}(u_i)$. Towards solving the non-cooperative game, the concept of Nash equilibrium is adopted. The Nash equilibrium (NE) of the non-cooperative game G is the vector of users' amount of offloaded data $\mathbf{b}^* = [b_1^*, \ldots, b_i^*, \ldots, b_N^*]$, where no user has the incentive to change its own strategy (i.e., amount of offloaded data) given the strategies of the rest of the users. Let $\mathbf{b}_{-\mathbf{i}}^* = [b_1^*, \ldots, b_{i-1}^*, b_{i+1}^*, \ldots, b_N^*]$ denote the vector of offloaded data of all users except user i at the NE point.

Definition 1: The offloaded data vector to the MEC server $\mathbf{b}^* = [b_1^*, \ldots, b_i^*, \ldots, b_N^*] \in S, S = S_1 \times \cdots \times S_N$ is a Pure Nash Equilibrium (PNE) point of the non-cooperative game $G = [\aleph, \{S_i\}, \{\mathbb{E}(u_i)\}]$, if $\forall i \in \aleph$ it holds true that $\mathbb{E}(u_i(b_i^*, \mathbf{b}_{-\mathbf{i}}^*)) \geq \mathbb{E}(u_i(b_i, \mathbf{b}_{-\mathbf{i}}^*)), \forall b_i \in S_i$.

It is noted that a Prospect Theory based game-theoretic approach is adopted to treat the aforementioned problem, instead of other approaches, due to its distributed and computationally efficient nature, while properly capturing the users' behavioral characteristics. The sequential best response dynamics mechanism is adopted to determine the game's PNE, which as also confirmed later in the paper, converges fast to it and in a scalable manner, due to its best response nature, in contrast for example to other learning based techniques, which need large exploration and exploitation time to determine a stable solution. In addition, in several cases a large amount of reliable data and extensive time would be required for the proper training of of a supervised learning based approach, for instance. However, a machine learning (ML) based approach (and in particular reinforcement leaning) could further complement the proposed framework and support the applicability of a best response determining process, in terms of treating potential incompleteness of the available information under uncertain environments, such as the communication and computing environment.

B. EXISTENCE AND UNIQUENESS OF PNE

Let us denote the best response strategy $B_i(\mathbf{b}_{-\mathbf{i}}) : S_{-i} \rightrightarrows S_i$ of user i, as follows:

$$B_i(\mathbf{b}_{-\mathbf{i}}) = \arg\max_{b_i \in S_i} \mathbb{E}(u_i(b_i, \mathbf{b}_{-\mathbf{i}})), \mathbf{b}_{-\mathbf{i}} \in S_{-i} \tag{15}$$

where $\mathbf{b}_{-\mathbf{i}} = [b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_N]$ is the data vector of all users excluding user i, and $S_{-i} = S_1 \times \cdots \times S_{i-1} \times S_{i+1} \times \cdots \times S_N$ the corresponding mixed strategy.

Theorem 1: For each user $i, i \in \aleph$, its best response strategy exists and it is single-valued, such that $b_i^* = B_i(\mathbf{b}_{-\mathbf{i}}) = \arg\max_{b_i \in S_i} \mathbb{E}(u_i(b_i, \mathbf{b}_{-\mathbf{i}}))$.

We adopt the notation $b_{-i} = \sum_{j=1, j\neq i}^{N} b_j$ to depict the total offloaded data of all users except user $i, i \in \aleph$. In order to prove the above theorem, we first present Berge's Theorem [44] and then we prove the following Lemmas 1 - 3.

Theorem 2: Let Θ and X be two metric spaces, and $\Gamma : \Theta \rightrightarrows X$ a compact valued correspondence. Let the function $\Phi : X \times \Theta \to \mathbb{R}$ be jointly continuous in X and Θ. We define:

1) $\sigma(\theta) := \arg\max_{x \in \Gamma(\theta)} \Phi(x, \theta)$
2) $\Phi^*(\theta) := \max_{x \in \Gamma(\theta)} \Phi(x, \theta), \forall \theta \in \Theta$

If Γ is continuous at $\theta \in \Theta$, then

1) $\sigma : \Theta \rightrightarrows X$ is compact-valued, upper hemicontinuous and closed at θ
2) $\Phi^* : \Theta \to \mathbb{R}$ is continuous at θ

Lemma 1: For each user $i, i \in \aleph$ the following hold true:

1) $b_i^* = 0$ if and only if $b_{-i} \geq \overline{b_i}$, where a value $\overline{b_i} \in [0, b_{th}]$ exists.
2) $b_i^* > 0$ and $b_i^* + b_{-i} < \overline{b_i}$, if $b_{-i} < \overline{b_i}$ and there exists an interval $A_i \subset [0, \overline{b_i})$ such that $g_i(\overline{b_i}) = 0$.

Proof: Initially, we clarify that the user's $i, i \in \aleph$ best response strategy b_i^* can either be zero, i.e., $B_i(\mathbf{b}_{-\mathbf{i}}) = b_i^* = 0$ or a positive value, i.e., $B_i(\mathbf{b}_{-\mathbf{i}}) = b_i^* \in S_i$, and the best response value can never be equal to b_{th}, i.e., $b_i^* = b_{th}$ as in this case $p(b_{th}) = 1$, thus the MEC server (CPR) fails and the user's expected prospect-theoretic utility is negative. The first order derivative of the effective rate of return of the MEC server for the user $i, i \in \aleph$ is given as follows:

$$\frac{\partial g_i(b_T)}{\partial b_T} = \frac{\partial \overline{d}_i(b_T)}{\partial b_T}(1 - p(b_T)) - \frac{\partial p(b_T)}{\partial b_T}(\overline{d}_i(b_T) + k_i\epsilon_i) \tag{16}$$

It is obvious that $\frac{\partial p(b_T)}{\partial b_T} > 0$, as the probability of failure, i.e., $p(b_T)$ is strictly increasing with respect to b_T. Also $(1 - p(b_T)) > 0$. Moreover, $\frac{\partial \overline{d}_i(b_T)}{\partial b_T} < 0$, since the $\overline{d}_i(b_T)$ is strictly decreasing with respect to b_T based on Proposition 1. Thus, $g_i(b_T)$ is strictly decreasing with respect to b_T.

CASE A: If $g_i(0) \leq 0$, then $g_i(b_i) \leq 0, \forall b_i \in S_i$ and $\mathbb{E}(u_i) \leq 0$. So, in this case the only best response for the user i is the zero value, i.e., $B_i(\mathbf{b}_{-\mathbf{i}}) = b_i^* = 0$. As a result, $\overline{b_i} = 0$ and the interval A_i is not defined.

CASE B: If $g_i(0) > 0$, then since $g_i(b_{th}) = -k_i\epsilon_i < 0$, we know that $\exists \overline{b_i} \in [0, b_{th}]$ such that $g_i(\overline{b_i}) = 0$ based on the Intermediate Value Theorem [45]. As a result, if $b_{-i} \geq \overline{b_i}$ then $B_i(\mathbf{b}_{-\mathbf{i}}) = 0$, as $\forall b_i \neq 0$, it holds true that $g_i(b_i + \overline{b_i}) < 0$ due to the fact that $g_i(\cdot)$ is strictly decreasing, thus $\mathbb{E}(u_i(b_i, \mathbf{b}_{-\mathbf{i}})) < 0$. On the other hand, if $b_{-i} < \overline{b_i}$ then $g_i(b_i + b_{-i}) > 0, \forall b_i \in (0, \overline{b_i} - b_{-i})$. So, in this case $\exists b_i : g_i(b_i + b_{-i}) > 0$, thus, the zero value cannot be the best response for the user $i, i \in \aleph$, if and only if $b_{-i} \in [0, \overline{b_i})$. Also, because of the positive value of the expected prospect-theoretic utility at the best response, i.e., $\mathbb{E}(u_i(b_i^*, \mathbf{b}_{-\mathbf{i}})) > 0$ it is true that $b_i^* + b_{-i} \in (0, \overline{b_i})$, and as a result the interval A_i exists and is defined as $A_i = (0, \overline{b_i})$ ■

Lemma 2: For each mobile user $i, i \in \aleph$, its best response b_i^* is single-valued $\forall b_{-i} \in S_{-i}$

Proof: Based on Lemma 1, case A, we have shown that the best response strategy is single-valued, i.e., $B_i(\mathbf{b}_{-\mathbf{i}}) = 0$ if and only if there exists a value $\overline{b_i} \in [0, b_{th}]$ such that $b_{-i} \geq \overline{b_i}$. Thus, in the following we examine the case B as presented in Lemma 1, where we have already shown that there exists at least one best response strategy, i.e., $B_i(\mathbf{b}_{-\mathbf{i}}) > 0$. Given that there exists at least one best response strategy $B_i(\mathbf{b}_{-\mathbf{i}})$, it should be one of the solutions of the expected prospect-theoretic utility's first order derivative, as follows:

$$\begin{aligned}\frac{\partial \mathbb{E}(u_i)}{\partial b_i} &= [b_i^{a_i}\frac{\partial \overline{d}_i(b_T)}{\partial b_T} + a_i b_i^{a_i - 1}\overline{d}_i(b_T)](1 - p(b_T)) \\ &\quad - b_i^{a_i}\overline{d}_i(b_T)\frac{\partial p(b_T)}{\partial b_T} \\ &\quad - k_i\epsilon_i[a_i b_i^{a_i-1}p(b_T) + b_i^{a_i}\frac{\partial p(b_T)}{\partial b_T}]\end{aligned} \tag{17}$$

It is noted that $\frac{\partial \mathbb{E}(u_i)}{\partial b_i} = \frac{\partial \mathbb{E}(u_i)}{\partial b_T}$, since $b_T = b_i + b_{-i}$. Also, $-b_i^{a_i}\overline{d}_i(b_T)\frac{\partial p(b_T)}{\partial b_T} < 0$ and $-k_i[a_i b_i^{a_i-1}p(b_T)+b_i^{a_i}\frac{\partial p(b_T)}{\partial b_T}] < 0$. Thus, to determine the root of Eq. 17, it should hold true:

$$[b_i^{a_i}\frac{\partial \overline{d}_i(b_T)}{\partial b_T} + a_i b_i^{a_i-1}\overline{d}_i(b_T)] > 0 \tag{18}$$

Calculating the second derivative of $\mathbb{E}(u_i)$ we have:

$$\begin{aligned}\frac{\partial^2 \mathbb{E}(u_i)}{\partial b_i^2} &= [b_i^{a_i}\frac{\partial^2 \overline{d}_i(b_T)}{\partial b_T^2} + 2a_i b_i^{a_i-1}\frac{\partial \overline{d}_i(b_T)}{\partial b_T}](1-p(b_T))\\ &\quad - 2b_i^{a_i-1}[b_i\frac{\partial \overline{d}_i(b_T)}{\partial b_T} + a_i\overline{d}_i(b_T)]\frac{\partial p(b_T)}{\partial b_T}\\ &\quad - b_i^{a_i}\overline{d}_i(b_T)\frac{\partial^2 p(b_T)}{\partial b_T^2}\\ &\quad - k_i\epsilon_i[2a_i b_i^{a_i-1}\frac{\partial p(b_T)}{\partial b_T} + b_i^{a_i}\frac{\partial^2 p(b_T)}{\partial b_T^2}]\\ &\quad + a_i(a_i-1)b_i^{a_i-2}[\overline{d}_i(b_T)(1-p(b_T)) - k_i\epsilon_i p(b_T)]\end{aligned} \tag{19}$$

Specifically, due to the fact that b_i satisfies (18), $\frac{\partial^2 \overline{d}_i(b_T)}{\partial b_T^2} < 0$, $\frac{\partial \overline{d}_i(b_T)}{\partial b_T} < 0$, $\frac{\partial p(b_T)}{\partial b_T} > 0$ and $\frac{\partial^2 p(b_T)}{\partial b_T^2} = 0$, it is true that $\frac{\partial^2 \mathbb{E}(u_i)}{\partial b_i^2} < 0, \forall b_i \in (0, \overline{b}_i)$, thus $\mathbb{E}(u_i)$ is strictly concave.

Moreover, given that $\overline{d}_i(b_T)$ is concave decreasing, the function from the inequality (18), i.e., $b_i^{a_i}\frac{\partial \overline{d}_i(b_T)}{\partial b_T} + a_i b_i^{a_i-1}\overline{d}_i(b_T)$, is decreasing with respect to b_i. For small values of b_i, i.e., $b_i \to 0$ and $b_{-i} < \overline{b_i}$ it holds true that $b_i^{a_i}\frac{\partial \overline{d}_i(b_T)}{\partial b_T} + a_i b_i^{a_i-1}\overline{d}_i(b_T) > 0$. Defining $C := \sup\{b_i \in S_i : b_i^{a_i}\frac{\partial \overline{d}_i(b_T)}{\partial b_T} + a_i b_i^{a_i-1}\overline{d}_i(b_T) > 0\}$, inequality 18 holds true only in the interval $[0, C]$. Thus, the expected prospect-theoretic utility function has a unique maximum in $[0, C]$. ■

Lemma 3: The best response strategy of the user $i, i \in \aleph$, $b_i^* : S_{-i} \rightrightarrows S_i$ is continuous for $b_{-i} \in S_{-i}$.

Proof: The $b_i^* : S_{-i} \rightrightarrows S_i$ is mapped to σ and the expected prospect-theoretic utility is mapped to the function Φ (see the notation in Theorem 2). We compute $b_i^* \in S_i$ and define the correspondence $\Gamma : S_{-i} \rightrightarrows [0, 1]$ for any joint strategies of users other than i. Therefore, Γ is compact valued, and both upper and lower hemicontinuous. Hence, b_i^* is upper hemicontinuous from Theorem 2 and as it is single-valued (Lemma 2), is continuous. ■

Based on Theorem 2 and Lemmas 1 - 3, we proved that for each user i, its best-response strategy $B_i(\mathbf{b_{-i}})$ exists and is single-valued and continuous. Thus, we proved Theorem 1.

Theorem 3: A pure Nash equilibrium $\mathbf{b}^* = [b_1^*, \ldots, b_N^*]$ of the non-cooperative game $G = [\aleph, \{S_i\}, \{\mathbb{E}(u_i)\}]$ exists.

Proof: The strategy set $S_i, \forall i \in \aleph$ is a convex compact subset of the Euclidean space and so is the joint strategy space, $S = S_1 \times \cdots \times S_N \subset \mathbb{R}^{|N|}$. By defining a mapping $T : S \to S$ such that $T(b_1, \ldots, b_N) = (b_1^*, \ldots, b_N^*)$, from Lemma 2, T is single-valued and from Lemma 3 is continuous. Brouwer's fixed point theorem guarantees the existence of a strategy profile $s = \{b_i^*\}_{i\in\aleph} \in S$ that is invariant under the best response mapping and therefore is a PNE of G [44]. ■

The best response b_i^* of user $i, i \in \aleph$ satisfies the equation $\frac{\partial \mathbb{E}(u_i)}{\partial b_i}|_{b_i^*} = 0$. Based on the latter condition and Eq. 13, we define the function $h_i(b_T) = \frac{-a_i g_i(b_T)}{\frac{\partial g_i(b_T)}{\partial b_T}} = b_i$ which satisfies $h_i(b_i^* + b_{-i}) = b_i^*$, when $b_i^* > 0$.

Lemma 4: The function $h_i(b_T)$ is strictly decreasing with respect to b_T, $b_T \in A_i$, where A_i is as defined in Lemma 1.

Proof: We have that $\overline{d}_i(b_T)$ is decreasing, and that:

$$\frac{1}{a_i}\frac{\partial h_i(b_T)}{\partial b_T} = -\frac{(\frac{\partial g_i(b_T)}{\partial b_T})^2 - g_i(b_T)\frac{\partial^2 g_i(b_T)}{\partial b_T^2}}{(\frac{\partial g_i(b_T)}{\partial b_T})^2} \tag{20}$$

The numerator is equal to $[\frac{\partial \overline{d}_i(b_T)}{\partial b_T}(1 - p(b_T))]^2 + [\frac{\partial p(b_T)}{\partial b_T}(\overline{d}_i(b_T) + k_i\epsilon_i)]^2 - g_i(b_T)\rho - 2\overline{d}_i(b_T)\frac{\partial p(b_T)}{\partial b_T}k_i\epsilon_i$, which is positive, and $\rho = \frac{\partial^2 \overline{d}_i(b_T)}{\partial b_T^2}(1 - p(b_T)) - \overline{d}_i(b_T)\frac{\partial^2 p(b_T)}{\partial b_T^2} - k_i\epsilon_i\frac{\partial^2 p(b_T)}{\partial b_T^2} \leq 0$. Thus, $h_i(b_T)$ is strictly decreasing. ■

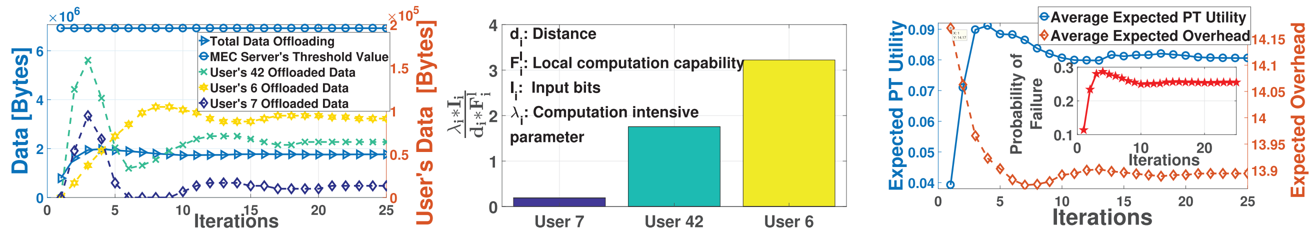
Theorem 4: The pure Nash Equilibrium of the non-cooperative game $G = [\aleph, \{S_i\}, \{\mathbb{E}(u_i)\}]$ is unique.

Proof: We use the notation b_T^* to denote the total amount of offloaded data at the PNE point of the game G. The proof of Theorem 4 is based on the reduction to absurdity. Let us suppose that we have two distinct PNE points, $b_{T(1)}^*, b_{T(2)}^*$. Without loss of generality we assume that $b_{T(2)}^* > b_{T(1)}^*$. We define the set $Sup \triangleq \{i \in \aleph : b_T^* < \overline{b}_i\}$, thus it includes every user with non zero amount of offloading data to the MEC server. Thus, we have $Sup_2 \subseteq Sup_1$. We have that $\sum_{j\in Sup_1} h_j(b_{T(1)}^*) = b_{T(1)}^*$, $\sum_{j\in Sup_2} h_j(b_{T(2)}^*) = b_{T(2)}^*$. So, $\sum_{j\in Sup_2} h_j(b_{T(1)}^*) + \sum_{j\in Sup_1 \setminus Sup_2} h_j(b_{T(1)}^*) = b_{T(1)}^* \Rightarrow \sum_{j\in Sup_2} h_j(b_{T(1)}^*) \leq b_{T(1)}^* < b_{T(2)}^* = \sum_{j\in Sup_2} h_j(b_{T(2)}^*)$. However, $h_i(b_T)$ is decreasing, so $h_i(b_{T(1)}^*) > h_i(b_{T(2)}^*), \forall j \in Sup_2$, which is contradiction. So, $b_{T(1)}^* = b_{T(2)}^*$. ■

C. ALGORITHM - CONVERGENCE TO PNE

A direct consequence of Lemma 4, is that the best response strategy of a user is decreasing in the total amount of offloading data. As a result, G belongs to the class of *best-response potential games*, thus, the sequential best response dynamics converge to the PNE of the game G [46]. From Theorem 4 and Lemma 2, we conclude that each user's i best response is unique. Specifically, its best response is zero if and only if the total offloaded data of the rest users is greater than its threshold value, i.e., $b_{-i} \geq \overline{b}_i$. Otherwise, its best response must be the root of the first order derivative of the expected prospect-theoretic utility, thus $\frac{\partial \mathbb{E}(u_i)}{\partial b_i} = 0$.

Each user i in order to compute its best response, first receives the total amount of offloaded data of the rest users, i.e., b_{-i} and then determines if it is zero, i.e., whether $b_{-i} \geq \overline{b_i}$ holds true. The later is satisfied if and only if $g_i(b_{-i}) \leq 0$. If the user i finds that $b_{-i} < \overline{b}_i$, then the best response b_i^* exists and is single-valued (Theorem 1). Specifically, due to the existence of the unique root of $\frac{\partial \mathbb{E}(u_i)}{\partial b_i} = 0$, and regarding



(a) Users' optimal offloaded data vs iterations

(b) Users' factor $\frac{\lambda_i * I_i}{d_i * F_i^l}$

(c) Average expected overhead and PT utility

FIGURE 3. **Pure operation of the proposed framework.**

that $\frac{\partial^2 \mathbb{E}(u_i)}{\partial b_i^2} < 0$, thus $\frac{\partial \mathbb{E}(u_i)}{\partial b_i}$ is strictly decreasing with respect to b_i, then the unique root r_i^* can be found via binary search into $[0, b_{th}]$ with an approximation ϵ, such that $\epsilon \rightarrow 0$, and finally the best response to be $b_i^* = min(I_i, r_i^*)$.

The complexity of the binary search is $\mathcal{O}(\log_2\ b_{th})$ [47]. In each iteration, N users execute Algorithm 1 and given that the rest operations involve arithmetical calculations and *Ite* iterations are needed for convergence to the PNE, the complexity of the distributed Algorithm 1 is $\mathcal{O}(N * Ite * \log_2\ b_{th})$. It is noted that the iterations scale very well with respect to the increasing number of users (see Section V-B).

Algorithm 1 Distributed Algorithm for Convergence to PNE

Input: *Set of Users* $\aleph = \{1, 2, \ldots, i, \ldots, N\}$
Output: Vector at PNE $\mathbf{b}^* = [b_1^*, \ldots, b_i^*, \ldots, b_N^*]$
$Ite = 0, b_i = randi(0, min(b_{th}, I_i)), \forall i \in \aleph$
while *PNE not reached* **do**
  $Ite = Ite + 1$
  for $i = 1$ to N **do**
    User i receives the vector $\mathbf{b_{-i}}$
    if $(g_i(b_{-i}) \leq 0)$ **then**
      $b_i^* = 0$
    else
      $r_i^* = BinarySearch([0, b_{th}], \epsilon)$
      $b_i^* = min(I_i, r_i^*)$
    end if
  end for
  Check convergence to PNE
end while

V. NUMERICAL RESULTS

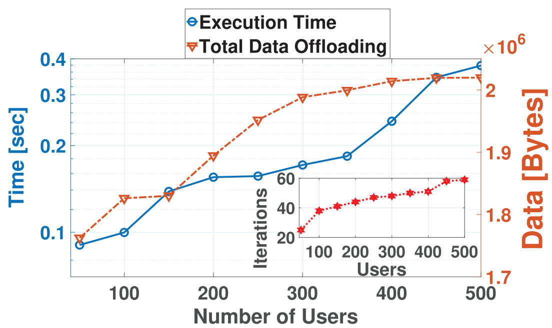
In this section, we provide some numerical results illustrating the operation, features and benefits of the proposed prospect-theoretic framework. In Section V-A, we focus on the pure operational characteristics of our framework, in terms of efficiently controlling the devices' offloaded data. A scalability analysis is provided in Section V-B, while in Section V-C we study our framework's operation under heterogeneous devices. Finally, in Section V-D, a comparative evaluation of our approach against alternative offloading strategies is provided.

In our study, we consider a base station acting as a MEC server with a coverage area of radius $R_0 = 50$m and $N = 50$ devices. Device's channel gain is modeled as $g_i = \frac{1}{d_i^{\theta}}$, where d_i is the distance of device i, $i \in \aleph$ from the MEC server, and θ is the distance loss exponent (e.g., $\theta = 2$). The transmission's bandwidth is considered $W = 5$MHz. Each device transmits with power $P_i = \frac{d_i^2}{R_0^2} Watt$, which is proportional to its distance from the MEC server. For each device we consider $F_i^l \in [0.1, 1]$ GHz and $f_i = 10^{-9} \frac{Joules}{Cpu-Cycle}, \forall i \in \aleph$ [48]. A face recognition application is considered with $I_i \in [1000, 2000]$ KB and $C_i \in [1000, 2000]$ Mega-Cycles [20], [21]. In the following, unless otherwise explicitly stated, we assume homogeneous users with parameters $a_i = 0.2$ and $k_i = 5$, $\forall i \in \aleph$. For the MEC server we consider that $F_{MEC} = 10^3$ GHz and $b_{th} = 10\% \times \sum_{i=1}^{50} I_i = 6.92 \times 10^6\ Bytes$. The proposed framework's evaluation was conducted via modeling and simulation in the MATLAB 9.7 R2019b simulation environment and executed in a MacBook Pro Laptop, 2.5GHz Intel Core i7, with 16GB LPDDR3 available RAM.

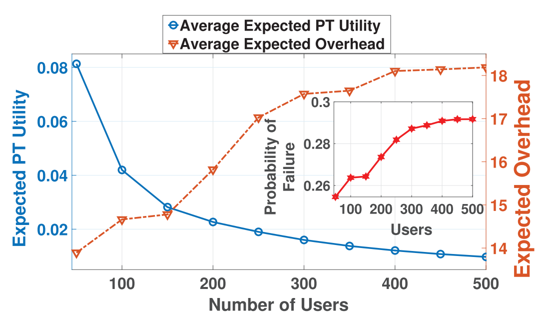
A. PURE OPERATION OF THE ALGORITHM

Fig. 3a presents for three indicative users the optimal amount of offloaded data to the MEC server, as a function of the iterations required to converge to the PNE point. We observe that for practical purposes less than twenty iterations are required to reach the PNE, starting from randomly selected initial values of offloaded data. Moreover, each device converges to a different amount of offloaded data, as its decision-making is based both on the MEC server's congestion, and its characteristics. Devices' characteristics are better captured by the factor $\frac{\lambda_i * I_i}{d_i * F_i^l}$, presented in Fig. 3b, which indicates that if a device has either high computation capability (F_i^l) or long distance (d_i) from the server, then it desires to offload a lower amount of data to avoid the transmission's overhead. However, the more demanding is the device's application (i.e., increased value of $\lambda_i * I_i$) the more the offloading action is desired by the device, so as to reduce the local overhead by executing part of its application at the server. Thus, the higher is the factor $\frac{\lambda_i * I_i}{d_i * F_i^l}$, the more beneficial is for the device to offload a larger amount of data to the MEC server.

Fig. 3c illustrates the average expected overhead and prospect-theoretic (PT) utility as a function of the iterations, while the MEC server's probability of failure is shown in the contained sub-figure. The results reveal that initially the expected average overhead and prospect-theoretic utility, are decreasing and increasing respectively, while the probability of failure also increases, as initially the MEC server is not congested (low probability of failure) due to the initial random feasible values of users' offloaded data. Thus, the users have high incentive to increase their offloaded data to reduce their overhead and increase their expected prospect-theoretic utility. As the time evolves, this trend leads to the MEC server's overloading (increased probability of failure) and the offloading action becomes less beneficial. Therefore, after a certain point the total offloading at the MEC server reduces and its corresponding probability of failure also reduces, while the users' offloaded data converge to the PNE point, i.e., $b_i^*, \forall i \in \aleph$ (Fig. 3a).



(a) Execution time and total offloaded data vs no. of users



(b) Average expected overhead and PT utility vs no. of users

FIGURE 4. Scalability evaluation.

B. SCALABILITY EVALUATION

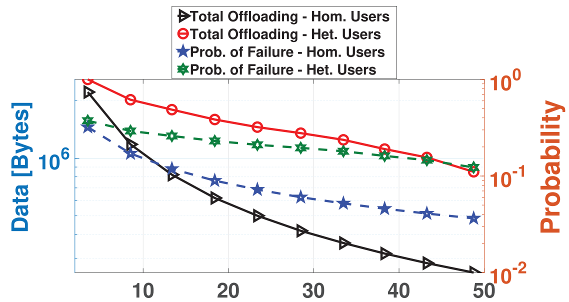
Fig. 4a illustrates the necessary time for convergence to PNE (and as contained sub-figure the corresponding iterations), as well as the total amount of offloaded data to the MEC server, as the number of users increases. It is observed that our prospect-theoretic framework scales very well with respect to the increasing number of users, as for an almost ten-fold increase in the number of users, the execution time increases at a significantly lower rate, i.e., four-fold increase. Similar observations follow with respect to the actual number of iterations as well. Furthermore, Fig. 4b presents the average expected overhead and prospect-theoretic utility at the PNE with respect to the increasing number of users. As the competition for MEC server's computing increases (i.e., increased number of users), the MEC server becomes more congested, and this results to a higher probability of failure, as it is depicted by the contained sub-figure. In this case, the offloading action for each user becomes less beneficial, and thus each user offloads reduced amount of data while executing a bigger portion of its task locally.

C. HETEROGENEOUS DEVICES - LOSS AVERSION

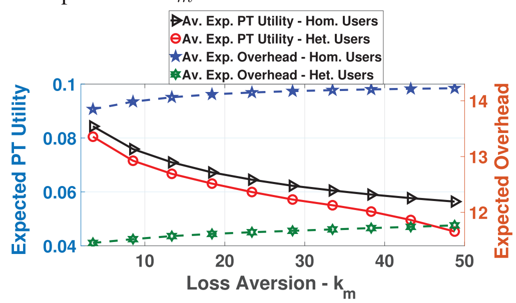
In this section, the impact of the users' heterogeneous loss aversion prospect-theoretic behavior on the the achievable performance is studied. Specifically, the results presented in Fig. 5a and Fig. 5b compare a scenario of homogeneous users (i.e., same loss aversion parameter k_m for all users) against a heterogeneous scenario, where each user i, $i \in \aleph$, is associated with a different personalized loss aversion index k_i. For a fair comparison, for all the users of the homogeneous group the considered loss aversion parameter k_m is equal to the average loss aversion parameter value of all the members of the heterogeneous group. That is, $k_m = \frac{\sum_{i=1}^{N} k_i}{N}$. The results reveal that the heterogeneous environment lead to higher congestion levels of the MEC server, as both the total amount of offloaded data and the MEC server's probability of failure, reach higher values (Fig. 5a), when compared to the corresponding ones of the homogeneous scenario. Furthermore, it is observed that the expected prospect-theoretic utility (Eq. 13) is decreasing with respect to the total amount of offloaded data b_T, and as expected the case of the heterogeneous users achieves lower average expected prospect-theoretic utility (Fig. 5b), due to the higher congestion levels of the MEC server. Furthermore, from Fig. 5b we note that the heterogeneous users, by offloading greater amount of their tasks to the MEC server, experience lower average expected overhead.

D. COMPARATIVE ANALYSIS

Considering the basic setting of homogeneous devices, a comparative study of the proposed optimal approach and solution demonstrates its superiority and benefits over alternative strategies. The comparative evaluation is performed with respect to the following metrics: achievable average expected overhead, MEC server's probability of failure, and total amount of offloaded data. Specifically, we compare our approach, which assumes prospect-theoretic users (PT), to three different approaches, that assume the following users' behaviors: (a) overhead minimizers (OM) users, who selfishly select their offloaded data in order to minimize their expected overhead, (b) only offloading (OO) users, who are totally risk seeking and offload their whole task to the MEC server, and (c) only local (OL) computing users, who are risk averse and keep the task execution locally, in order to obtain the "safe" and guaranteed performance provided by their own devices.



(a) Total offloaded data and probability of failure vs loss aversion parameter k_m



(b) Average expected overhead and PT utility vs loss aversion parameter k_m

FIGURE 5. Heterogeneous users - loss aversion impact study.

TABLE 2. Comparative evaluation.

User's Nature	Average Expected Overhead	Probability of Failure	Total Offloaded Data [Bytes]
PT	13.8927	0.2546	1762475
OM	14.2385	0.9438	6534778
OO	14.8586	1	76932000
OL	14.3116	0	0

Table 2 summarizes the corresponding results. Based on the fourth column of Table 2, we confirm that the OL users (last row) do not offload any data to the MEC server as expected, the OO users offload all of their data, the OM users offload a significant (but not the whole) amount of their data aiming at minimizing their overahed, while the PT users consider the server's probability of failure and accordingly offload a moderate amount of data. Specifically, when we consider totally risk seeking (OO) or risk averse (OL) users, they experience the worst overhead, as either they lead the MEC server to failure with probability 1 (case of OO users), or they do not exploit its high computation capability by not offloading any part of their task (case of OL users), respectively. Also, the overhead of the OO users is greater than the one of the OL users, as the first ones have an extra overhead owing to their transmissions. On the other hand, with respect to the case of the OM users, even though lower overhead is achieved compared to the previous cases since the users inherently aim at neutral overhead minimization, the selfish users' behavior does not consider the MEC server's failure probability, and eventually leads the MEC server to overloaded status and high probability of failure. Finally, the PT users achieve the lowest average expected overhead compared to all the other approaches, while at the same time MEC server's probability of failure remains at significant lower values.

VI. CONCLUSIONS AND FURTHER RESEARCH

In this paper, a device-centric risk-based distributed approach was proposed to determine the users' IoT devices' computation offloading volume in a wireless MEC environment, taking into consideration the loss averse and gain seeking behavior of the users in accordance with the properties of Prospect Theory. The proposed model and approach, is enabled by and enables cognition, and comes in contrast to the majority of existing methods in the literature, that adopt the expected utility maximization theory, where the users are assumed as risk-neutral. In our setting the MEC server acts as a common pool of resources with uncertain payoff returned to the devices, due to its shared nature and the corresponding interdependence among the users' devices, while the choice of computation executed at the local IoT device was assumed to be a safe computation option. Exploiting the local IoT device computing and total offloading overhead, while taking into account each user's cognitive biases and behavior, the optimal amount of each user's offloaded data to the MEC server was obtained as the outcome of a non-cooperative game among the users. The existence and uniqueness of a PNE point was shown, and an algorithm that converges to the optimal values of offloaded data for each user in a distributed manner, was designed. Detailed numerical results were obtained via modeling and simulation, that demonstrated the operation features and superiority of the proposed cognitive-enabled framework, under both cases of homogeneous and heterogeneous users.

It should be noted that in our current work we have assumed the existence of a single MEC server for potential offloading. Part of our current and future work, focuses on the adoption and extension of the above approach in an environment where multiple MEC servers co-exist acting as potential common pool of resources, and treat the problem of both horizontal and vertical offloading scaling and performance in this environment, while still capturing user risk perceptions.

Furthermore, it is noted that given the inherent uncertainty in the communication and computing environment under consideration, an open issue becoming of high practical and research importance, is the investigation of how different forms of the considered MEC server probability of failure function, can capture different operational factors and characteristics of the server and its usage (e.g., server's robustness to failure, reliability, various non-linear server's behaviors to traffic loads, soft vs. hard failures etc.).

Finally, it is noted that in the emerging era of Internet of Things and Tactile Internet, the requirements for ultra low-latency and energy efficiency, are becoming critical and challenging issues for several applications. In our work the latency and energy issues have been directly considered as part of the corresponding overheads computed for both offloading and local computing, while they are essentially treated through the optimization of these overheads. However, an interesting extension of our modeling, could be to further consider hard latency and energy constraints for some applications as part of the overall optimization approach.

REFERENCES

[1] CV Networking, "Internet of Things at a glance," Cisco, San Jose, CA, USA, White Paper, 2017.

[2] Ericsson, "Ericsson mobility report," Ericsson, Stockholm, Sweden, Ericsson Mobility Rep., Nov. 2019.

[3] R. Dautov, S. Distefano, D. Bruneo, F. Longo, G. Merlino, and A. Puliafito, "Data processing in Cyber-Physical-Social systems through edge computing," *IEEE Access*, vol. 6, pp. 29822–29835, 2018.

[4] N. Hassan, S. Gillani, E. Ahmed, I. Yaqoob, and M. Imran, "The role of edge computing in Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 11, pp. 110–115, Nov. 2018.

[5] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 2017.

[6] S. Zanafi, N. Aknin, M. Giacobbe, M. Scarpa, and A. Puliafito, "Enabling sustainable smart environments using fog computing," in *Proc. Int. Conf. Electron., Control, Optim. Comput. Sci. (ICECOCS)*, Dec. 2018, pp. 1–6.

[7] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795–2808, Oct. 2016.

[8] Y. Wang, "Special issue on cognitive computing, on abstract intelligence," *Int. J. Softw. Sci. Comput. Intell.*, vol. 1, no. 3, p. 2, 2009.

[9] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behav. Brain Sci.*, vol. 40, 2017, Art. no. e253.

[10] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Joint allocation of computation and communication resources in multiuser mobile cloud computing," in *Proc. IEEE 14th Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2013, pp. 26–30.

[11] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5994–6009, Sep. 2017.

[12] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Power-delay tradeoff in multi-user mobile-edge computing systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[13] O. Munoz, A. Pascual-Iserte, and J. Vidal, "Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4738–4755, Oct. 2015.

[14] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.

[15] Y. Yu, J. Zhang, and K. B. Letaief, "Joint subcarrier and CPU time allocation for mobile edge computing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[16] K. Zhang, Y. Mao, S. Leng, Q. Zhao, L. Li, X. Peng, L. Pan, S. Maharjan, and Y. Zhang, "Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks," *IEEE Access*, vol. 4, pp. 5896–5907, 2016.

[17] H. Yu, Q. Wang, and S. Guo, "Energy-efficient task offloading and resource scheduling for mobile edge computing," in *Proc. IEEE Int. Conf. Netw., Archit. Storage (NAS)*, Oct. 2018, pp. 1–4.

[18] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4268–4282, Aug. 2016.

[19] S. Guo, B. Xiao, Y. Yang, and Y. Yang, "Energy-efficient dynamic offloading and resource scheduling in mobile cloud computing," in *Proc. IEEE 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[20] M.-H. Chen, B. Liang, and M. Dong, "Joint offloading and resource allocation for computation and communication in mobile cloud with computing access point," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[21] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 974–983, Apr. 2015.

[22] S. Josilo and G. Dan, "A game theoretic analysis of selfish mobile computation offloading," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.

[23] K. Zhang, Y. Mao, S. Leng, S. Maharjan, and Y. Zhang, "Optimal delay constrained offloading for vehicular edge computing networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[24] K. Xie, X. Wang, G. Xie, D. Xie, J. Cao, Y. Ji, and J. Wen, "Distributed multi-dimensional pricing for efficient application offloading in mobile cloud computing," *IEEE Trans. Services Comput.*, vol. 12, no. 6, pp. 925–940, Nov. 2019.

[25] S. Ranadheera, S. Maghsudi, and E. Hossain, "Mobile edge computation offloading using game theory and reinforcement learning," 2017, *arXiv:1711.09012*. [Online]. Available: http://arxiv.org/abs/1711.09012

[26] Y. Tao, Y. Zhang, and Y. Ji, "Efficient computation offloading strategies for mobile cloud computing," in *Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl.*, Mar. 2015, pp. 626–633.

[27] C.-F. Liu, M. Bennis, and H. V. Poor, "Latency and reliability-aware task offloading and resource allocation for mobile edge computing," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2017, pp. 1–7.

[28] A. U. R. Khan, M. Othman, A. N. Khan, J. Shuja, and S. Mustafa, "Computation offloading cost estimation in mobile cloud application models," *Wireless Pers. Commun.*, vol. 97, no. 3, pp. 4897–4920, Dec. 2017.

[29] J. Shuja, A. Gani, K. Ko, K. So, S. Mustafa, S. A. Madani, and M. K. Khan, "SIMDOM: A framework for SIMD instruction translation and offloading in heterogeneous mobile architectures," *Trans. Emerg. Telecommun. Technol.*, vol. 29, no. 4, Apr. 2018, Art. no. e3174.

[30] M. Lewandowski, "Prospect theory versus expected utility theory: Assumptions, predictions, intuition and modelling of risk attitudes," *Central Eur. J. Econ. Model. Econ.*, vol. 9, no. 4, pp. 275–321, 2017.

[31] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *J. Risk Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[32] L. Tang and S. He, "Multi-user computation offloading in mobile edge computing: A behavioral perspective," *IEEE Netw.*, vol. 32, no. 1, pp. 48–53, Jan. 2018.

[33] A. Thanou, E. E. Tsiropoulou, and S. Papavassiliou, "Quality of experience under a prospect theoretic perspective: A cultural heritage space use case," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 1, pp. 135–148, Feb. 2019.

[34] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "Dynamic spectrum management in 5G wireless networks: A real-life modeling approach," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2019, pp. 2134–2142.

[35] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "On controlling spectrum fragility via resource pricing in 5G wireless networks," *IEEE Netw. Lett.*, vol. 1, no. 3, pp. 111–115, Sep. 2019.

[36] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "On the prospect of UAV-assisted communications paradigm in public safety networks," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 762–767.

[37] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "Exploiting prospect theory and risk-awareness to protect UAV-assisted network operation," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 286, Dec. 2019.

[38] L. Xiao, J. Liu, Y. Li, N. B. Mandayam, and H. V. Poor, "Prospect theoretic analysis of anti-jamming communications in cognitive radio networks," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 746–751.

[39] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 291–303, 2002.

[40] G. Hardin, "The tragedy of the commons," *J. Natural Resour. Policy Res.*, vol. 1, no. 3, pp. 243–253, 2009.

[41] A. R. Hota, S. Garg, and S. Sundaram, "Fragility of the commons under prospect-theoretic risk attitudes," *Games Econ. Behav.*, vol. 98, pp. 135–164, Jul. 2016.

[42] N. Barberis, M. Huang, and T. Santos, "Prospect theory and asset prices," *Quart. J. Econ.*, vol. 116, no. 1, pp. 1–53, Feb. 2001.

[43] G. P. Fettweis, "The tactile Internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

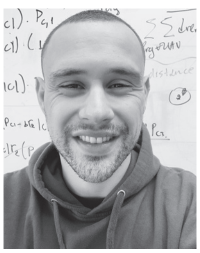
[44] E. A. Ok, *Real Analysis With Economic Applications*, vol. 10. Princeton, NJ, USA: Princeton Univ. Press, 2007.

[45] S. B. Russ, "A translation of Bolzano's paper on the intermediate value theorem," *Historia Math.*, vol. 7, no. 2, pp. 156–185, May 1980.

[46] P. Dubey, O. Haimanko, and A. Zapechelnyuk, "Strategic complements and substitutes, and potential games," *Games Econ. Behav.*, vol. 54, no. 1, pp. 77–94, Jan. 2006.

[47] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2009.

[48] A. P. Miettinen and J. K. Nurminen, "Energy efficiency of mobile clients in cloud computing," in *Proc. HotCloud*, vol. 10, 2010, p. 4.



PAVLOS ATHANASIOS APOSTOLOPOULOS (Student Member, IEEE) received the Diploma degree in electrical and computer engineering from the National Technical University of Athens, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of New Mexico. He is also a Research Assistant with the University of New Mexico. His main research interests include machine learning, deep learning, game theory, behavioral decision theory, and optimization.

EIRINI ELENI TSIROPOULOU (Member, IEEE) is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of New Mexico. Her main research interests are in the area of cyber-physical social systems and wireless heterogeneous networks, with emphasis on network modeling and optimization, resource orchestration in interdependent systems, reinforcement learning, game theory, network economics, and the Internet of Things. Three of her articles received the Best Paper Award at the IEEE WCNC, in 2012, ADHOCNETS, in 2015, and the IEEE/IFIP WMNC 2019. She was selected by the IEEE Communication Society—N2Women—as one of the top ten Rising Stars of 2017 in the communications and networking field.

SYMEON PAPAVASSILIOU (Senior Member, IEEE) was a Senior Technical Staff Member with AT&T Laboratories, Middletown, NJ, USA, from 1995 to 1999. In August 1999, he joined the ECE Department, New Jersey Institute of Technology, USA, where he was an Associate Professor, until 2004. He is currently a Professor with the School of ECE, National Technical University of Athens. He has an established record of publications in his field of expertise, with more than 300 technical journals and conference published papers. His main research interests are in the area of computer communication and computing networks, with emphasis on the analysis, optimization, and performance evaluation of mobile and distributed systems, wireless networks, mobile edge computing architectures, stochastic, and complex systems. He received the Best Paper Award in the IEEE INFOCOM 94, the AT&T Division Recognition and Achievement Award, in 1997, the U.S. National Science Foundation Career Award, in 2003, the Best Paper Award in the IEEE WCNC 2012, the Excellence in Research Grant, Greece, in 2012, and the best paper awards in ADHOCNETS 2015, ICT 2016, and the IEEE/IFIP WMNC 2019. He has also served on the board of the Greek National Regulatory Authority on Telecommunications and Posts, from 2006 to 2009.