A chance-constraint approach for optimizing social engagement-based services

Michel Bierlaire*, Edoardo Fadda†, Lohic Fotio Tiotsop‡, and Daniele Manerba§

*Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne - CH-1015 EPFL, Lausanne, (Switzerland)
Email: michel.bierlaire@epfl.ch

†Department of Mathematical Sciences, Politecnico di Torino - corso Duca degli Abruzzi 24, Torino (Italy)
Email: edoardo.fadda@polito.it

‡Department of Control and Computer Engineering, Politecnico di Torino - corso Duca degli Abruzzi 24, Torino (Italy)
Email lohic.fotiotiotsop@polito.it

§Department of Information Engineering, Università degli Studi di Brescia, via Branze 38, Brescia (Italy)
Email: daniele.manerba@unibs.it

***Abstract*—Social engagement is a novel business model whose goal is transforming final users of a service from passive components into active ones. In this framework, people are contacted by the decision-maker (generally a company) and they are asked to perform tasks in exchange for a reward. This paves the way to the interesting optimization problem of allocating the different types of workforce so as to minimize costs. Despite this problem has been investigated within the operations research community, there is no model that allows to solve it by explicitly and appropriately modeling the behavior of contacted candidates through consolidated concepts from utility theory. This work aims at filling this gap. We propose a stochastic optimization model including a chance constraint that puts in relation, under probabilistic terms, the candidate *willingness to accept* a task and the reward actually offered by the decision-maker. The proposed model aims at optimally deciding which user to contact, the amount of the reward proposed, and how many employees to use in order to minimize the total expected costs of the operations. A solution approach is proposed to address the formulated stochastic optimization problem and its computational efficiency and effectiveness are investigated through an extensive set of computational experiments.**

I. Introduction

The social engagement business models is a new business paradigm involving the customers of a company in its operations. More precisely, people agree to perform specific services for a company in exchange for a reward. The social engagement business model has been enabled by the increase of the number of users connected on the web and technologies able to get people information [1]. This gives to the companies the possibility to easily communicate with customers and other people, and then to propose tasks in exchange for a reward.

A concrete example of application of the social engagement paradigm is the so called *crowd-shipping* logistics, i.e., rather than building an entire network of transshipment facilities and deploying a large number of vehicles for the delivery, the companies ask the people to collect the packages to a certain location and deliver it to the final user [2]. By using crowd-shipping the company does not only decrease its costs, but also decreases its environmental impact. In fact, people accepting the delivery usually would take advantage of travels that they have to do anyhow for other activities, thus avoiding additional trips.

Another really interesting application of social engagement occurs in the development of the innovative concept called Internet of Things (IoT). In the era of smart cities, IoT is playing a role of considerable importance [3]. However, its development is considerably slowed down by the difficulty and costs involved in building telecommunication networks capable of continuously transmitting large amounts of data collected by sensors, in an efficient, effective and reliable way. As an alternative to the construction of this network, in exchange for a reward, citizens use their devices (e.g., mobile phones, modems) to share the internet so that the nearby sensors can exploit it to communicate the gathered data. This is known as the opportunistic IoT (oIoT) paradigm [4].

In the light of these examples, it is clear that social engagement will acquire more and more importance in the future being a corner stone of both smart cities and sustainable behaviour and that a great variety of applications will soon be put in practice. In this work, therefore, we do not want to concentrate on a specific application rather on a very general social engagement-based setting in order to embrace all the basic characteristics of such a business model. In the rest of the paper, we will call a *candidate* each person that might be contacted by the company for operation proposal, and a *task* each proposed operation.

An effective planning of operations under the social engagement paradigm yields an interesting optimization problem. The decision-maker must decide how much he is willing to pay to a candidate for each task, when and where to rely on employees and on candidates, which tasks to assign to the employees and for which tasks the candidates must be contacted, in order to minimize the total operational costs. It is important to note that the reward paid to a candidate is generally lower on average than the cost that the company bears for an employee. However, while an employee is obliged to accept and carry out the tasks assigned to him, there is no certainty that a candidate will accept a proposed task. This complicates the underlying optimization problem as one must correctly address this source of uncertainty.

A little attention has been devoted to the development of optimization models aimed at effectively scheduling companies operations that exploit social engagement. Just few works ([5], [6], [7]) have tried to tackle the problem and, therefore, there is a large room for improvement of existing approaches as well as for the design of more innovative and more complete ones (as claimed regarding crowd-shipping in [8]). In particular, to the best of our knowledge, there is no published optimization model that explicitly accounts for individual candidate behaviour when planning social engagement-based operations. As already mentioned, one characteristic that makes challenging the optimization problems deriving from the implementation of the social engagement paradigm is the fact that candidates are not constrained a priori to respect a contract. This means that, once contacted, the candidate may not accept the task and, if we assume a pure rational profit-maximization behavior of the candidate, the reject can happen because the proposed reward is lower than the candidate expectation. It is therefore important to integrate tools in the decision-making process that allow monitoring the individual behavior of potential candidates.

In this work, to account for individual behaviour, we rely on the candidate's *willingness to accept* a task, i.e., the minimum reward expected by a candidate to accept a task. The *willingness to accept* is a well consolidated concept in utility theory and has been used since long to explain human subject preferences in economics [9]. From the decision-maker point of view, the candidate's *willingness to accept* is not deterministically known, since it depends on some factors that are intrinsic of the candidates. Therefore, we consider the candidate *willingness to accept* as a random variable. Thus, the probability of acceptance for a candidate will be equal to the probability that the offered reward is greater or equal to the *willingness to accept* of the candidate. Given this knowledge, the aim of the decision maker is to decide which person to contact, the amount of the reward proposed, and how many employees to use in order to minimize the total operations cost composed expressed as sum between the costs of the employee and the expected reward for the candidates.

This paper's contribution is threefold and can be summarized as it follows:

- First, we propose a novel mathematical model for social engagement-based services optimization. The formulation, which includes probability constraints, results to be the first one that explicitly accounts for each individual candidates behaviour.
- Second, since the complexity of the proposed model and the explicit consideration of stochastic parameters do not allow to obtain a simple solution, we derive a mixed-integer quadratic programming model that approximates the original model. This is done by making some reasonable hypothesis on the probability distribution of the *willingness to accept* of each candidate, and by exploiting the Markov inequality.
- Third, we conduct several computational experiment to validate and assess the suitability of our proposed model and of the solution approach and their characteristics.

The rest of the paper is organized as it follows. Section II discusses the related works appeared in the literature. The considered optimization problem is defined in Section III, while the related mathematical model is formulated in Section IV. Our solution approach is described in Section V. Section VI presents the computational experiments settings, while the related results are discussed in Section VII. Fianlly, Section VIII draws conclusions and highlights future developments.

II. Related Works

Operation research models and methods have long been leveraged for cost reduction or profit maximization by companies. With the advent of the social engagement business model, the operation research community has naturally focused on the question of how to effectively plan operations that rely on such a business model.

As pointed out in the introduction, one of the most interesting implementation of the social engagement business model is the crowd-shipping paradigm. A recent survey on operation research approaches to crowd-sourced delivery can be found in [8]. Among the open research directions listed, one regards how to determine the minimum value of the reward guaranteeing that each candidate accepts the proposed task. In fact, the decision-maker does not know a priori how many candidates will accept the proposed tasks. As a result, some papers as [10] assume that the candidates will surely accept the task proposed. However, such an assumption yields strong limitations in practice [8]. Others authors, e.g., [11], [6] have proposed models that better fit practical requirements by considering the possibility that a contacted candidate may reject the tasks. In [6], the authors assume that the number of person willing to perform a task in a given area of the operational network is a random variable. In this paper, we adopt a similar perspective. However, instead of relying on a single random variable describing the number of candidates, we model each single candidate behavior through a Bernoulli random variable. The parameter of such a Bernoulli random variable, i.e. the probability that the candidate accepts the task, is not fixed but depends by the proposed reward.

To deal with any single candidate's behaviour, we consider the concept of *willingness to accept* from the utility theory [12]. This is actuality not a peculiarity of this paper since many authors have relied on utility theory concepts and tools to model individual behaviours in optimization problems. In [13] an approach to embed discrete choice models from the utility theory into mixed integer linear programming models was presented. Instead, this paper wants to be a first step toward integrating utility theory tools and concepts also in optimization models for social engagement-based services, so to appropriately model individual candidates behaviors.

In our optimization problem, the link between the offered reward, the minimum expectations of a candidate, and his/her probability to accept the task is expressed under probabilistic terms, naturally yielding an optimization model including chance constraints. Optimization problems involving chance constraints have been largely considered and studied in literature [14]. Although some approaches have been proposed to deal with these problems [15], [16], there is a common agreement that they are in general too complex to solve directly and thus the design of ad-hoc solution approaches might be required.

III. The social engagement optimization problem

The social engagement optimization problem that we want to study considers a decision-maker (in general a company) whose goal is to use people, in the following called *candidate*, in addition to employees in order to perform a set of tasks. In particular, we consider a urban environment divided in several geographical areas such as mobile phone cells, neighborhoods of different market or just geographical areas. Each of these areas is characterized by a number of tasks to perform and each tasks is characterized by different workloads, thus a single task may requires more candidates to be done. This is frequently the case in oIoT applications in which the tasks can be associated to share the internet connection with more sensors in the same area, thus requiring more candidates.

For example, in the crowd shipping setting these tasks are the delivery required by customers out of the store, while in the oIoT application these tasks consist in sharing the internet connection with smart sensors in the city.

Each task can either be performed by using an employee or a candidate. Employee are more expensive, are available in a small number but they execute the tasks assigned. Instead, candidates are less expensive, their quantity is virtually unlimited (since the number of people considered for social engagement is far greater than the number of tasks) but they can refuse to perform a task with a given probability. We assume that the acceptance probability increases as the offered reward increase. Please notice that the employee have greater productivity than the candidates.

The goal of the decision-maker is to minimize the total operative costs while enforcing that with high probability all the tasks must be performed.

In the next sections, we will consider the following sets:

- $\mathcal{I}$: set of tasks.
- $\mathcal{M}$: set of candidates.

Moreover, we define the following parameters:

- c_i: cost of using an employee for task i.
- W_i: workload required to perform task i.
- $r > 1$: ratio between the productivity of an employee with respect to that of a candidate, i.e., the workload that a single employee can afford as compared to a candidate in the same time frame.
- α_i: required probability for tasks i to be performed.
- Δ_i^m: random variable representing the *willingness to accept* of candidates m for task i.
- B: maximum number of available employees.

IV. Mathematical Model

In this section we formulate the mathematical model for the social engagement optimization problem. We consider the following decision variables:

- $Q_i^m \in \mathbb{R}$: reward offered to candidate m to accept task i;
- $z_i \in \mathbb{N}$: number of employee assigned to tasks i;

and the following auxiliary variables:

- $x_i^m \in [0, 1]$: probability for candidate m to accept task i.
- Y_i^m are a random variables distributed according to a Bernoulli distribution of probability x_i^m, i.e.,

$$Y_i^m = \begin{cases} 1 \text{ if candidate } m \text{ accepts to perform task } i \\ 0 \text{ otherwise} \end{cases}.$$

Thus the Social Engagement Optimization Problem ($SEOP$) can be formulated as follows:

$$\min \sum_{i\in\mathcal{I}} \sum_{m\in\mathcal{M}} Q_i^m x_i^m + \sum_{i\in\mathcal{I}} c_i z_i \tag{1}$$

subject to

$$x_i^m = \mathbb{P}[Q_i^m \geq \Delta_i^m], \quad i \in \mathcal{I}, m \in \mathcal{M} \tag{2}$$

$$\mathbb{P}[Y_i^m = a] = (x_i^m)^a (1 - x_i^m)^{(1-a)}, \quad i \in \mathcal{I}, m \in \mathcal{M} \tag{3}$$

$$a \in \{0, 1\} \tag{4}$$

$$\mathbb{P}\left[\sum_{m\in\mathcal{M}} Y_i^m + r z_i \geq W_i\right] \geq \alpha_i, \quad i \in \mathcal{I} \tag{5}$$

$$\sum_{i\in\mathcal{I}} z_i \leq B \tag{6}$$

$$\begin{aligned} & z_i \in \mathbb{N}, \quad i \in \mathcal{I}, \\ & Q_i^m \in \mathbb{R}^+, x_i^m, Y_i^m \in [0, 1], \quad i \in \mathcal{I}, m \in \mathcal{M}. \end{aligned} \tag{7}$$

The objective function (1) is the total cost expressed as the summation between the expected costs offered as a reward (the reward Q_i^m is paid with probability x_i^m), and the sum of the costs from the employee. Constraints (2) define the variables x_i^m as the acceptance probability, while constraints (3) and (4) ensure Y_i^m to follow a Bernoulli distribution. Constraints (5) are chance constraints enforcing a minimum probability of doing a given task either by using employee or candidates. It is worth noting that ensuring that each task is satisfied with a given probability is less strict than requiring that all the tasks will be satisfied with a given probability. Nevertheless enforcing this second condition would lead to too conservative solutions. Finally, constraint (6) limits the number of employees.

In conclusion, it is worth noting that the $SEOP$ model is quite general, embracing all the possible applications based on a social engagement business model.

A. An analytical solution for the single-candidate $SEOP$

It is worth noting that it is reasonable to apply social engagement if $|\mathcal{M}|$ is large. However, in the particular case where there is a single candidate it is possible to derive an analytical expression of the optimal solution. In this section we derive such an expression since it provides some preliminary insights into how the candidate's behavior is related to the decision-maker's expectations in terms of how likely he performs the task. In particular, we consider $|\mathcal{M}| = 1|$, and for the sake of simplicity we also assume that:

- just one task requiring a unitary effort has to be performed, i.e. $|\mathcal{I}| = 1$, $W_1 = 1$
- each employee can perform just one task, namely $r = 1$.

The resulting model is

$$\min_{z,Q} QF_\Delta(Q) + cz \tag{8}$$

subject to

$$\mathbb{P}[Y = a] = (F_\Delta(Q))^a(1 - F_\Delta(Q))^{(1-a)} \tag{9}$$

$$a \in \{0, 1\} \tag{10}$$

$$\mathbb{P}[Y \geq 1 - z] \geq \alpha \tag{11}$$

$$z \in \{0, 1\}, Q \in \mathbb{R}^+, x \in [0, 1]. \tag{12}$$

where F_Δ denotes the cumulative probability distribution (cdf) of the *willingness to accept* (i.e. Δ). To obtain Model (8)–(12) from $SEOP$, we drop the index i for ease of notation, we rearrange the terms in constraint (11), we set $z \in \{0, 1\}$ due to the characteristic of the instance and we set $x = F_\Delta(Q)$ due to constraint (2). Since Y in Eq. (11) is a Bernoulli random variable and the $1 - z \in \{0, 1\}$, we have that

$$\begin{cases} \text{if } z = 1, \ \ \mathbb{P}[Y \geq 0] \geq \alpha \implies 1 \geq \alpha \\ \text{if } z = 0, \ \ \mathbb{P}[Y \geq 1] \geq \alpha \implies F_\Delta(Q) \ \geq \alpha. \end{cases}$$

Hence, if $z = 1$ constraint (11) is always satisfied, else it must hold that $F_\Delta(Q) \ \geq \alpha$ or, equivalently that $Q \geq F_\Delta^{-1}(\alpha)$. Since the goal of the problem is to minimize the total costs, the optimal solution will consider $Q = F_\Delta^{-1}(\alpha)$. Hence, the optimal solution to the problem can be obtained by considering the relation between the cost c (the cost paid to the employee) and $\alpha F_\Delta^{-1}(\alpha)$ (the expected costs for the candidate's reward). In particular, if $c \leq \alpha F_\Delta^{-1}(\alpha)$, the optimal solution is $z = 1, Q = 0$ with value c, otherwise it is $z = 0, Q = F_\Delta^{-1}(\alpha)$, with value of $\alpha F_\Delta^{-1}(\alpha)$.

In other words, in this simple example the company uses social engagement if $\alpha F_\Delta^{-1}(\alpha) \ \leq \ c$. Hence, the higher is α, the more convenient are the employees since $F_\Delta^{-1}(\alpha)$ is increasing. This claim is also verified in the real field in which social engagement is used for tasks that are not safety critical.

Unfortunately, the insights obtained for $SEOP$ in the single-candidate case cannot be generalized in more complex scenarios involving more candidates, since the chance constraint is coupling the related variables. In fact, considering two candidates ($|M| = 2$), a task requiring $W_1 = 2$, and considering two employee ($B = 2$), we have that

$$\begin{cases} \text{if } z = 0 \ \ \mathbb{P}[Y_1 + Y_2 \geq 2] \geq \alpha \implies F_{\Delta_1}(Q^1)F_{\Delta_2}(Q^2) \geq \alpha \\ \text{if } z = 1 \ \ \mathbb{P}[Y_1 + Y_2 \geq 1] \geq \alpha \implies \\ F_{\Delta_1}(Q^1)(1 - F_{\Delta_2}(Q^2)) + (1 - F_{\Delta_1}(Q^1))F_{\Delta_2}(Q^2) \geq \alpha \\ \text{if } z = 2 \ \ \mathbb{P}[Y \geq 0] \geq \alpha \implies 1 \geq \alpha. \end{cases}$$

This means that, even for $|\mathcal{M}| = 2$, we lose an easy interpretation and, in turn, an easy solution for the problem.

V. Solution Approach

The optimization problem (1)-(6) is difficult to solve due to the definition of the x_i^m in constraints (2), of Y_i^m in constraints (3) and (4), and of the chance constraints in constraint (5). Hence, we approximate these constraints in order to get a model which can be readily solved with off-the-shelf solvers.

Constraints (2) involve the cdf of the random variable Δ_i^m. We approximate it by means of a piece-wise linear function with J breakpoints. In particular, instead of constraints (2) we add a set of constraints of the form

$$x_i^m \leq k_j Q_i^m + q_j, \quad j = 1, \ldots, J, i \in \mathcal{I}, m \in \mathcal{M}, \tag{13}$$

where k_j and q_j are obtained by imposing proper conditions (e.g. the passage in J points of the cdf). This choice is equivalent to enforce $x_i^m \leq \min[1, m_1 Q_i^m + q_1, \ldots, m_J Q_i^m + q_J]$, where the first term of the minimum comes form the x_i^m definition. Since the approximation proposed in Eq. (14) just lead to concave functions (being the pointwise minimum of affine functions) and since the a general cdf may be convex in some portion of the domain, the proposed approximation is not guarantee to converge to the cdf for all the distributions. In the following, for the sake of simplicity, we consider just $J = 1$ and we impose the passage for the point $(0, 0)$ meaning that with 0 reward the probability that the candidate will perform the task is 0 and $(\bar{Q}_i^m, 1)$ where $\bar{Q}_i^m$ is a quantity of reward such the candidate is willing to perform the task with a probability that we may approximate to be 1. By making this choice, the obtained approximation is:

$$x_i^m \leq Q_i^m / \bar{Q}_i^m, \quad i \in \mathcal{I}, m \in \mathcal{M}. \tag{14}$$

This choice leads to the approximation of the cdf depicted in red in Figure 1. Considering Eq. (14) is equivalent to suppose that Δ_i^m is distributed according to a uniform distribution. Thus, adding more functions is equivalent to approximate the density function with a function that is piece-wise constant.

In order to deal with constraints (5) we notice that by splitting the random variable from the deterministic components, the probability that we aim to compute is

$$\mathbb{P}[\sum_{m \in \mathcal{M}} Y_i^m \geq W_i - rz_i] \geq \alpha_i, \quad i \in \mathcal{I}. \tag{15}$$

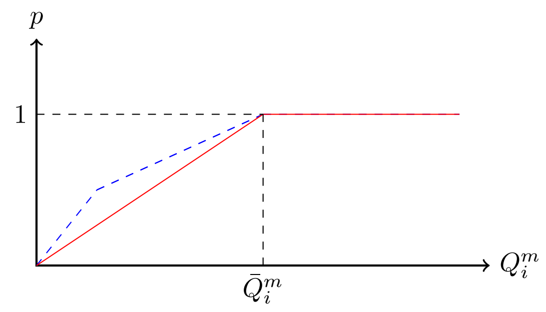

Fig. 1. Eq. (2) approximation In red simple piece-wise approximation, in dashed blue a more finer approximation.

By using Markov inequality (i.e., the fact that if X is a nonnegative random variable and $a > 0$, it holds that $\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$), we can write:

$$\frac{\mathbb{E}[\sum_{m\in\mathcal{M}} Y_i^m]}{W_i - rz_i} \geq \mathbb{P}\left[\sum_{m\in\mathcal{M}} Y_i^m \geq W_i - rz_i\right] \geq \alpha_i, \quad i \in \mathcal{I}. \tag{16}$$

Hence, since

$$\mathbb{E}\left[\sum_{m\in\mathcal{M}} Y_i^m\right] = \sum_{m\in\mathcal{M}} \mathbb{E}[Y_i^m] = \sum_{m\in\mathcal{M}} x_i^m,$$

Eq. (16) leads to the following constraint:

$$\sum_{m\in\mathcal{M}} x_i^m \geq \alpha_i(W_i - rz_i), \quad i \in \mathcal{I}. \tag{17}$$

Eq. (17) is enforcing that the expected workload form the candidates must be greater than the α_i percent of the people needed. Moreover, by considering the bound provided by Eq. (17), we are reducing the feasible set, thus the condition in (15) will be satisfied for greater value of α_i.

Then, the resulting approximation of the Social Engagement Optimization Problem ($SEOP_{ap}$) is the following model:

$$\min \sum_{i\in\mathcal{I}} \sum_{m\in\mathcal{M}} Q_i^m x_i^m + \sum_{i\in\mathcal{I}} c_i z_i \tag{18}$$

subject to

$$x_i^m \leq \frac{Q_i^m}{\bar{Q}_i^m}, \quad i \in \mathcal{I}, m \in \mathcal{M} \tag{19}$$

$$\sum_{m\in\mathcal{M}} x_i^m \geq \alpha_i(W_i - rz_i), \quad i \in \mathcal{I} \tag{20}$$

$$\sum_{i\in\mathcal{I}} z_i \leq B \tag{21}$$

$$\begin{aligned} &z_i \in \mathbb{N}, \quad i \in \mathcal{I}, \\ &Q_i^m \in \mathbb{R}^+, x_i^m \in [0,1], \quad i \in \mathcal{I}, m \in \mathcal{M}. \end{aligned} \tag{22}$$

$SEOP_{ap}$ is a mixed integer quadratic problem that can be solved in an exact way by using off the shelf solvers such as GUROBI.

VI. Experimental setting

In this section, we present some computational experiment considering $SEOP_{ap}$. Since space limitations preclude the description of a full-fledged experimental design, we prefer to give some glimpse in order to get a feeling for the problem and the impact of its features.

A. Data generation

Social engagement is a relatively new topic of research and very few data related to instances generation is available. Moreover, different applications are characterized by different magnitude of the parameters describing the instances. Since crowd-shipping and oIoT are the domains of application characterized by more data available, we start by some available data for these two settings to define general benchmark instances.

In crowd-shipping applications, the number of tasks $|\mathcal{I}|$ is generally considered between 5 and 20 and $W_i = 1, \forall i \in \mathcal{I}$, since each task is a single delivery [2]. Instead, in oIoT applications, it is reasonable to consider much greater number for $|\mathcal{I}|$ and W_i. Within a use case related to the collection of data from smart waste collection sensors [6], the authors consider $|\mathcal{I}| = 100$, with $W_i = 10, \forall i \in \mathcal{I}$. We, therefore, will consider $|\mathcal{I}| = \{5, 10, 20, 50, 100\}$.

The number of candidates is the most difficult parameter to tune since it may be arbitrarily large: we can assume it to be equal to the number of people that has agreed to receive information related to social engagement possibilities. Nevertheless, we suppose that to be $|\mathcal{M}| = 4|\mathcal{I}|$ because the goal of the paper is to present a model and its properties rather then solution methods able to solve big instances.

Concerning the parameters, we set $w_i = W * (0.8 + 0.4U)$, where W is proportional to the hourly wage, and U is a uniform random variable between $[0, 1]$. We set $W = 7$ € (being 14 € the average hourly wage for a non skilled worker) [17]. By making this assumption we are the implicit assuming that the time unit for the work effort will be a half an hour. In the crowd-shipping application this means that we assume that one employee will take half an hour to bring $r = 1$ parcels and to make deliveries, while in the oIoT we consider that to be the time for going in one area and to collect the data from $r = 10$ sensors. Given a number of task r that each employee can do, we consider the maximum number of employee to be $B = 20/r$ in the crowd-ship application and $B = 20/r$ in the oIoT one. With this choice, we assume that the instances have no feasibility problems being the employee enough to perform all the tasks.

The required probability for the task i to be performed is α_i. In general, we expect that to be different for all the tasks. For example, in crowd-shipping applications, an higher probability would be set for more important customers while in the oIoT application higher probability must be associated with sensors which information is more valuable. Since setting the α_i would be totally arbitrarily, we suppose that to be equal for all the tasks, i.e., $\alpha_1 = \alpha_2 = \cdots = \alpha_{|\mathcal{I}|} := \alpha$. Since the application is not safety critical and it is always possible to perform all the task with the employee, the values of α are considered in the interval $[0.6, 0.9]$.

Finally, Δ_i^m is the *willingness to accept*, i.e.,the minimum monetary quantity that a person is willing to accept to sell a good or service, or to bear a negative externality. In our case it is the minimum amount of money for which a candidate is willing to accept to make a task. While no study are available in the oIoT framework, in the crowd-shipping applications this quantity has been shown to depend by the size of the parcel, the sex of the person and by the time of the day [18], [19], [20]. Since modelling all these aspects require ad-hoc work, we postpone this topic in future study and we consider Δ_i^m to be a simple random variable. Since Δ_i^m can be assumed to be an unobserved utility, we consider it to be distributed according to a Gumbel (or Extreme Value distribution type I) [21]. The cdf of this distribution is: $F(s) = e^{-e^{-s}}$ and by twice differentiating it is possible to observe that it is a concave function, thus leading the approximation proposed in Section V to converge to the exact cdf. Another hypothesis done in utility theory is that unobserved utility are normally distributed, we left this assumption for future work since it does not have the same concave property of the Gumbel cdf. For calibrating the distribution, we consider the average reward for crowshipping operations. In particular, in [22] it is reported that reward are usually in the interval between 5-7 dollars equivalent to 4 to 6 €. Thus, we fix the average of the Gumbel to be equal to 5 €, with a standard deviation of 1 €. By using this distribution we then compute $\bar{Q}_i^m$ to be $\bar{Q}_i^m = F^{-1}_{\Delta_i^m}(0.99)$.

VII. Results and Analyses

In this section we report the computational experiments. All the experiments were performed on a *Intel(R) Core(TM)* i7-5500U CPU@2.40GHz computer with 16GB of RAM and running *Ubuntu* v20.04. The exact solver used is Gurobi v9.1.1 via its Python3 APIs.

A. Computational results

In this section we study the computational time with respect to the dimension of the problem of $SEOP_{ap}$. In particular, versus the growth of $|\mathcal{I}|$ and $|\mathcal{M}|$, we evaluate:

- the CPU time (sec).
- the time-to-best (sec): it is the number of seconds from the start of the execution of Gurobi to the time in which it founds the best solution of the run. This is an interesting parameter since often exact solvers find really fast the optimal solution, then they spend a lot of time for proving optimality. Thus, by observing this parameter is possible to better understand the performance of the exact solver.
- the MIP gap (%): it is computed as the percentage difference between the lower and upper objective bound. In particular, we consider the least gap value that Gurobi has to reach before stopping its execution.

The average and standard deviation on 10 instances are shown in Table I. In all the runs we set the solver time limit to 1 hour. We chose this threshold because the decision-maker may, in principle run the model several time a day and allowing for longer computational times will deepen its usability.

As the reader can notice, the instances with $|\mathcal{I}| = 5$, and $|\mathcal{M}| = 20$ are solved almost instantaneously with 0 gap. The time-to-best is equal to the CPU time since the difference are below the hundredths of a second.

For instances with $|\mathcal{I}| = 10, |\mathcal{M}| = 40$, and $|\mathcal{I}| = 20, |\mathcal{M}| = 80$, the computational time increases, but the solver is still able to find the optimal solution inside the time limit. For the instances with $|\mathcal{I}| = 10, |\mathcal{M}| = 40$ the time to best is near one half of the total computational time but solutions with gap below the 5% are found by the solver already in the first minutes of the run. Instead, for the instance with $|\mathcal{I}| = 20, |\mathcal{M}| = 80$, the time-to-best is close to the whole computation time and no solution with gap below the 5% is found in the first minute of the run. For instance of greater dimensions the exact solver is not able to find the optimal solution in the given time limit, for this reason the CPU time is equal to 3600 seconds with a standard deviation of 0. Nevertheless, for instances with $|\mathcal{I}| = 50$, and $|\mathcal{M}| = 200$, several times the final gaps are below the 10%, while for instances with $|\mathcal{I}| = 100$, and $|\mathcal{M}| = 400$, the problem is not able to find a good bound in the time limit, thus the 100 MIP gap with 0 standard deviation.

These results enable us to claim that $SEOP_{ap}$ can be used in the real setting for crowd-shipping applications due to the low number of delivery and people considered while for the oIoT setting heuristic solution method are required.

B. Approximation analysis

In Section V we propose an approximation of $SEOP$. In this section we analyse the goodness of the approximation by means of some computational experiment. Since Δ_i^m is distributed according to a Gumbel distribution with concave cdf, the approximation proposed converges to the exact function and several techniques for developing good piecewise approximation are available, e.g. the Concave Adaptive Value Estimation [23]. Thus, we are interested in how much conservative is the Markov inequality with respect to Eq. (5). In order to quantify this difference we compute the optimal solution of $SEOP_{ap}$ and we use it to compute $\hat{\alpha} := \mathbb{P}[\sum_{m \in \mathcal{M}} Y_i^m + rz_i \geq W_i]$. It is possible to easily compute this quantity by noting that the Y_i^m are independent with respect to the index m since the knowledge about candidate m performing a task does not provide any information related to the execution of the same task by other candidate. Thus, $\sum_m Y_i^m$ is a sum of independent random variable distributed according to Bernoulli distribution of parameter x_i^m. Central Limit Theorems for non identically distributed random variables are available and, in particular, by applying the Lyapunov Central Limit Theorem (see [24]) it is possible to prove that for large values of $|\mathcal{M}|$ (in practice $|\mathcal{M}| \geq 30$),

TABLE I

AVERAGE AND STANDARD DEVIATIONS OF THE CPU TIME, TIME TO BEST, AND MIP GAP FOR DIFFERENT VALUES OF $\mathcal{I}$, AND $\mathcal{M}$.

Instance		CPU time(sec)		time-to-best (sec)		MIP gap (%)	
$\|\mathcal{I}\|$	$\|\mathcal{M}\|$	Avg	Std Dev	Avg	Std Dev	Avg	Std Dev
5	20	0.09	0.01	0.09	0.01	0	0
10	40	1569.56	175.83	614.31	235.43	0	0
20	80	2780.84	573.26	2634.94	897.98	0	0
50	200	3600.00	0.00	1054.31	562.25	56	47
100	400	3600.00	0.00	1679.57	720.77	100	0

it holds that:

$$\sum_{m\in\mathcal{M}} Y_i^m \sim \mathcal{N}\left(\sum_{m\in\mathcal{M}} x_i^m, \sum_{m\in\mathcal{M}} x_i^m(1-x_i^m)\right), \quad i\in\mathcal{I}. \tag{23}$$

By using (23), we can compute α by solving:

$$\alpha_i = 1-\Phi\left(\frac{W_i - rz_i - \sum_{m\in\mathcal{M}} x_i^m}{\sqrt{\sum_{m\in\mathcal{M}} x_i^m(1-x_i^m)}}\right), \quad i\in\mathcal{I}, \tag{24}$$

where Φ is the cdf of a standard normal distribution. We report the value of the $\alpha \in [0.5, 1]$ in $SEOP_{ap}$ versus the $\hat{\alpha}$ computed with Eq. (24) in Figure 2. All the results are averaged over 10 runs and the standard deviation of the observation is represented as an uncertain area. We compute the results for $|\mathcal{I}| = 10$ and $|\mathcal{M}| = 40$ since we are able to get the optimal solution in a reasonable amount of time. Moreover, with $|\mathcal{M}| = 40$ there are enough candidates to let us apply results in Eqs. (23)–(24).

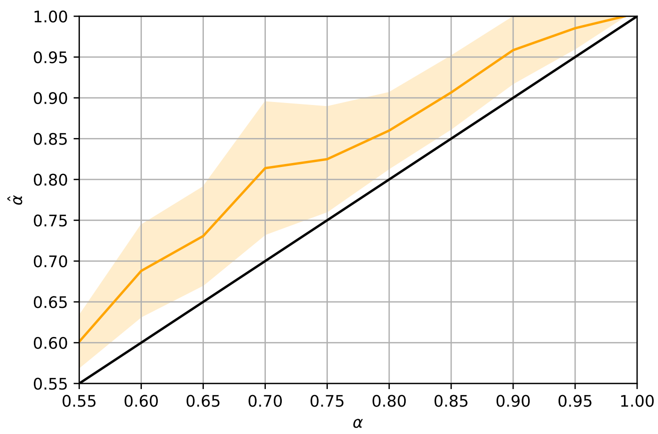

Fig. 2. Values of α set in the model vs real value of α.

As we expected, the curve is above the line $\hat{\alpha} = \alpha$ since by using the Markov inequality we are considering an upper bound on the probability. Nevertheless, the results are close to the exact value being on average 10% higher than the α set in the model. Thus, in the real field the decision-maker may lower by the 10% the values of the αs and get a solution compliant with the wanted probability of execution.

In conclusion, it is worth noting that using the result of Eq. (23) for Eq. (5) lead to the following equation

$$W_i - rz_i - \sum_{m\in\mathcal{M}} x_i^m \geq z_{1-\alpha}\sqrt{\sum_{m\in\mathcal{M}} x_i^m(1-x_i^m)}, \quad i\in\mathcal{I} \tag{25}$$

where $z_{1-\alpha}$ is the $1-\alpha$ standard normal quantile. Nevertheless, due to the negative square term on the right hand side of Eq. (25), it is not a second order cone constraint, preventing its usage to simplify $SEOP$.

C. Sensitivity analysis and managerial insights

Since for values of $|\mathcal{I}| = 10$ and $|\mathcal{M}| = 40$, the optimal solution can be computed in a reasonable amount of time, we study how the solution characteristics change in different settings. In particular, versus different values of α in the range $[0.5, 1]$, we will evaluate:

- the expected percentage costs for the employees' services, calculated as
$$\rho_E = \frac{\sum_{i\in\mathcal{I}} c_i z_i}{\sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} Q_i^m x_i^m + \sum_{i\in\mathcal{I}} c_i z_i};$$
- the percentage of services performed by employee, calculated as
$$\omega_E = r\frac{\sum_{i\in\mathcal{I}} z_i}{\sum_{i\in\mathcal{I}} W_i}.$$

The average results on 10 instances for ρ_E and ω_E are shown in Figure 3. Note that we do not show values of α lower than

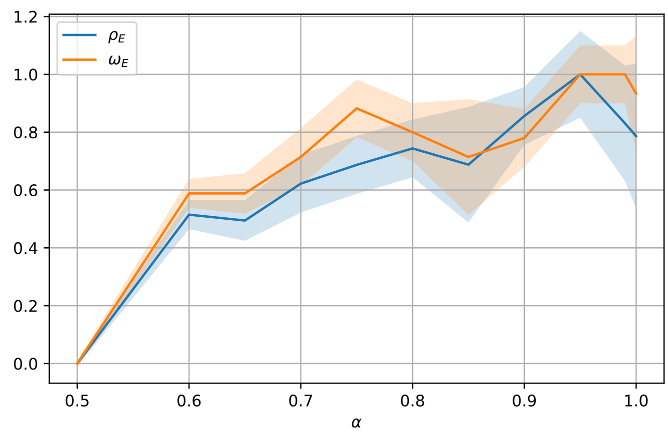

Fig. 3. Values of ρ_E and ω_E for different values of α.

0.5 since for those values the solution uses only candidates.

As α increases, the solutions use more the employee and less the candidates. It is interesting to notice that even for $\alpha = 1$, a small percentage of the candidates it is still selected. By analyzing the solution more in detail it is possible to see that the Q_i^m are set to several small values. Even if these values lead to small probability of acceptance x_i^m their summation satisfy constraint (20). This is an approximation error generated by the usage of the Markov Inequality. Nevertheless, it is not really important since the domain of usage of the model will consider lower values of α.

VIII. Conclusions and future works

In this study, we proposed a new probabilistic model for social engagement-based services optimization encompassing the *willingness to accept* of the candidate involved in the business model. We prove, by means of computational experiment that, despite the difficult formulation, the model can be approximated into a nice tractable form able to provide timely solution for crowd-shipping applications.

Being the social engagement a very seminal topic within the optimization field, several improvement can be sketched:

- first, a full-fledged experimental design to explore all the solution characteristics is needed. Some questions to answer are related to the performance of the method in the case in which non-concave distributions for the *willingness to accept* are considered or how the solutions of the model are related to the number of breakpoints used by the piece-wise linear cumulative distribution approximation of the *willingness to accept*.
- second, the problem paves the way to be faced by means of other techniques, such as dynamic programming, or other paradigms to treat the uncertainty, such as stochastic programming or robust optimization. In particular, we believe that another promising paradigm could be the distributionally robust optimization, since it relaxes the assumption of knowing the real characteristics of the willingness to accept distribution.
- finally, we highlight the need for ad-hoc methodologies in order to get solutions for oIoT applications (i.e., those having the larger number of tasks and candidates) in a reasonable time. This may be achieved in two ways: by introducing new approximation methodologies, or by developing heuristics for the proposed mixed integer non linear programming model.

We expect to cover some of these aspects in future studies.

References

[1] F. Corno, L. D. Russis, and T. Montanaro, "Estimate user meaningful places through low-energy mobile sensing," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, Oct. 2016. doi: 10.1109/smc.2016.7844703. [Online]. Available: https://doi.org/10.1109/smc.2016.7844703

[2] A. Santini, A. Viana, X. Klimentova, and J. P. Pedroso, "The probabilistic travelling salesman problem with crowdsourcing," *Computers & Operations Research*, vol. 142, p. 105722, Jun. 2022. doi: 10.1016/j.cor.2022.105722. [Online]. Available: https://doi.org/10.1016/j.cor.2022.105722

[3] A. Malagnino, T. Montanaro, M. Lazoi, I. Sergi, A. Corallo, and L. Patrono, "Building information modeling and internet of things integration for smart and sustainable environments: A review," *Journal of Cleaner Production*, vol. 312, p. 127716, Aug. 2021. doi: 10.1016/j.jclepro.2021.127716. [Online]. Available: https://doi.org/10.1016/j.jclepro.2021.127716

[4] B. Guo, D. Zhang, Z. Wang, Z. Yu, and X. Zhou, "Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things," *Journal of Network and Computer Applications*, vol. 36, no. 6, pp. 1531–1539, Nov. 2013. doi: 10.1016/j.jnca.2012.12.028. [Online]. Available: https://doi.org/10.1016/j.jnca.2012.12.028

[5] E. Fadda, G. Perboli, and R. Tadei, "A progressive hedging method for the optimization of social engagement and opportunistic IoT problems," *European Journal of Operational Research*, vol. 277, no. 2, pp. 643–652, Sep. 2019. doi: 10.1016/j.ejor.2019.02.052. [Online]. Available: https://doi.org/10.1016/j.ejor.2019.02.052

[6] ——, "Customized multi-period stochastic assignment problem for social engagement and opportunistic IoT," *Computers & Operations Research*, vol. 93, pp. 41–50, May 2018. doi: 10.1016/j.cor.2018.01.010. [Online]. Available: https://doi.org/10.1016/j.cor.2018.01.010

[7] E. Fadda, D. Mana, G. Perboli, and R. Tadei, "Multi period assignment problem for social engagement and opportunistic IoT," in *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*. IEEE, Jul. 2017. doi: 10.1109/compsac.2017.173. [Online]. Available: https://doi.org/10.1109/compsac.2017.173

[8] A. Alnaggar, F. Gzara, and J. H. Bookbinder, "Crowdsourced delivery: A review of platforms and academic literature," *Omega*, vol. 98, p. 102139, Jan. 2021. doi: 10.1016/j.omega.2019.102139. [Online]. Available: https://doi.org/10.1016/j.omega.2019.102139

[9] W. M. Hanemann, "Willingness to pay and willingness to accept: how much can they differ?" *The American Economic Review*, vol. 81, no. 3, pp. 635–647, 1991.

[10] C. Archetti, M. Savelsbergh, and M. G. Speranza, "The vehicle routing problem with occasional drivers," *European Journal of Operational Research*, vol. 254, no. 2, pp. 472–480, 2016.

[11] K. Gdowska, A. Viana, and J. P. Pedroso, "Stochastic last-mile delivery with crowdshipping," *Transportation research procedia*, vol. 30, pp. 90–100, 2018.

[12] P. C. Fishburn, "Utility theory," *Management science*, vol. 14, no. 5, pp. 335–378, 1968.

[13] M. P. Paneque, M. Bierlaire, B. Gendron, and S. S. Azadeh, "Integrating advanced discrete choice models in mixed integer linear optimization," *Transportation Research Part B: Methodological*, vol. 146, pp. 26–49, Apr. 2021. doi: 10.1016/j.trb.2021.02.003. [Online]. Available: https://doi.org/10.1016/j.trb.2021.02.003

[14] P. Li, H. Arellano-Garcia, and G. Wozny, "Chance constrained programming approach to process optimization under uncertainty," *Computers & chemical engineering*, vol. 32, no. 1-2, pp. 25–45, 2008.

[15] A. Nemirovski and A. Shapiro, "Scenario approximations of chance constraints," *Probabilistic and randomized methods for design under uncertainty*, pp. 3–47, 2006.

[16] K. Margellos, P. Goulart, and J. Lygeros, "On the road between robust optimization and the scenario approach for chance constrained optimization problems," *IEEE Transactions on Automatic Control*, vol. 59, no. 8, pp. 2258–2263, 2014.

[17] "Statistics explained." [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Wages_and_labour_costs

[18] A. Ermagun and A. Stathopoulos, "To bid or not to bid: An empirical study of the supply determinants of crowd-shipping," *Transportation Research Part A: Policy and Practice*, vol. 116, pp. 468–483, Oct. 2018. doi: 10.1016/j.tra.2018.06.019. [Online]. Available: https://doi.org/10.1016/j.tra.2018.06.019

[19] E. Marcucci, M. L. Pira, C. S. Carrocci, V. Gatta, and E. Pieralice, "Connected shared mobility for passengers and freight: Investigating the potential of crowdshipping in urban areas," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, Jun. 2017. doi: 10.1109/mtits.2017.8005629. [Online]. Available: https://doi.org/10.1109/mtits.2017.8005629

[20] J. Miller, Y. M. Nie, and A. Stathopoulos, "Crowdsourced urban package delivery," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2610, no. 1, pp. 67–75, Jan. 2017. doi: 10.3141/2610-08. [Online]. Available: https://doi.org/10.3141/2610-08

[21] D. McFadden, "The measurement of urban travel demand," *Journal of Public Economics*, vol. 3, no. 4, pp. 303–328, 1974. [Online]. Available: https://EconPapers.repec.org/RePEc:eee:pubeco:v:3:y:1974:i:4:p:303-328

[22] E. Pourrahmani and M. Jaller, "Crowdshipping in last mile deliveries: Operational challenges and research opportunities," *Socio-Economic Planning Sciences*, vol. 78, p. 101063, Dec. 2021. doi: 10.1016/j.seps.2021.101063. [Online]. Available: https://doi.org/10.1016/j.seps.2021.101063

[23] G. A. Godfrey and W. B. Powell, "An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution," *Management Science*, vol. 47, no. 8, pp. 1101–1112, Aug. 2001. doi: 10.1287/mnsc.47.8.1101.10231. [Online]. Available: https://doi.org/10.1287/mnsc.47.8.1101.10231

[24] A. Cuzzocrea, E. Fadda, and A. Baldo, "Lyapunov central limit theorem: Theoretical properties and applications in big-data-populated smart city settings," in *2021 5th International Conference on Cloud and Big Data Computing (ICCBDC)*. ACM, Aug. 2021. doi: 10.1145/3481646.3481652. [Online]. Available: https://doi.org/10.1145/3481646.3481652