Quantifying the Impact of Incentives on Service Availability at the Extreme Edge

Sherif B. Azmy[1], MHD Saria Allahham[2], Nizar Zorba[3], Hossam S. Hassanein[2]
[1]Department of Electrical and Computer Engineering, Queen's University, Kingston, ON, Canada
[2]School of Computing, Queen's University, Kingston, ON, Canada
[3]Electrical Engineering Department, Qatar University, Doha, Qatar
Email: {sherif.azmy,20msa7}@queensu.ca, nizarz@qu.edu.qa, hossam@cs.queensu.ca

GLOBECOM 2024 - 2024 IEEE Global Communications Conference | 979-8-3503-5125-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/GLOBECOM52923.2024.10901514

***Abstract*—Edge computing seeks to optimize service provision over enterprise-owned infrastructure near the end-user at the network's edge. However, it misses out on the opportunity to utilize user-owned hardware at the extreme edge of the network as workers in a sharing economy. In this work, we build upon the existing Incentive Vacation Queueing (IVQ) model and develop the Virtual Kiosk Model (VKM) to analyze service availability and the dynamics of multiple workers' participation in the provision of a service on the extreme edge. We formulate an optimization problem to minimize total cost of incentive payments while maintaining service availability under temporal constraints. We propose the Model-based Incentive Strategy at the Edge (MISE) algorithm to iteratively adjust incentives in real-time. MISE is compared against traditional numerical optimizers and a baseline naive approach that greedily focuses on minimizing incentives. Our findings demonstrate that MISE ensures sustained service availability without overburdening the workers at a cost acceptable to the service provider, striking a crucial balance in the management of extreme edge computing resources.**

I. Introduction

The transition from centralized cloud computing to edge computing represents a significant change in data processing paradigms by bringing computational resources closer to the end-users on the edge of the network. This distributive approach overcomes challenges in cloud computing regarding latency, bandwidth, and privacy [1]. Although edge computing primarily focused on enterprise-owned infrastructure, it misses out on a significant potential of user-owned resources on the extreme edge. Leveraging the wealth idle resources available on user-owned devices is potentially capable of transforming service availability, efficiency, and quality, as an extension to the reliable enterprise-owned edge computing [2].

Taking the edge to the very extreme is enabled thanks to recent technological developments. For instance, containerization and virtualization have radically transformed compute-provision over the last decade by utilizing containers instead of bulky virtual machines. Recently, even lighter microcontainers and unikernels were created to overcome the limitations of containers, allowing them to be deployed on resource-limited devices [3], [4]. In addition, coupled with 6G and beyond's communication capabilities, ad-hoc and distributed topologies on user-owned devices are becoming a promising direction that will enable a wide range of applications. In fact, as the demand increases for data-driven AI models, the ease of partitioning of such models allows running them in a distributed fashion; *exo* for example deploys an LLM on an AI cluster of personal devices [5].

User-owned edge resources are located within the end-user's proximity, which gives them an advantage to traditional edge in terms of latency. Devices such as smartphones, tablets, and even smart appliances, are not only idle, but they are also equipped with general-purpose hardware. However, exploiting such user-owned resources is laden with challenges due to the heterogeneity, limited capabilities, and sporadic usage nature of such devices [6], [7]. Opportunistically leveraging the resources of such personal devices requires establishing guarantees on reliability and availability, but this is challenging due to the influence of human behaviour [8]. However, by using incentives and allowing multi-tenancy, such unpredictable wild behaviour can be tamed. One way is by renting resources from user-owned workers in a manner similar to sharing economy systems such as Uber or Airbnb, but for computational infrastructure in what is called Reward Edge Computing (REC) [2], [9]. This approach allows access to resources at a cheaper cost while allowing these worker devices to generate income.

Similar notions predate REC and show that leveraging user-owned resources is potentially feasible. For instance, the authors in [8] analyze the performance of a peer-to-peer offloading network and provide a proof of concept that the user-owned devices can provide grid-like capabilities for some applications. In another work [10], crowdsourcing computation is explored with the intent to maximize social welfare for workers and service providers in a crowdsourcing market in which service providers incentivize workers. Economic incentives are influential when it comes to utilizing user-owned devices as they mitigate a wide range of issues. In [11], a resource auction trader system on the edge in which a service provider deploys a service on user-owned devices is developed. However, the literature is either scarce on the direct relationship between incentives and the performance of the system or rich on enterprise-owned edge computing.

In this work, we extend the initial model developed in [2] to quantify the impact of incentives on temporal service availability. A service provider seeks to keep a service available for their end-users, and as such they recruit a minimum number of workers to deploy the service in a distributed manner. We model such service provision a virtual kiosk model, which models

workers arrivals and departures as an availability sequence over time that extends to the system as a whole. However, guaranteeing availability does not ensure that costs are minimal or that the workers are not overworked. To strike a balance, we quantify the impact of incentives in the terms of the extra time gained if a lesser incentives had been paid. We then formulate an optimization problem to minimize the total cost under system availability and worker temporal constraints. However, snapshot optimization does not provide a significant performance improvement. To that end, we develop an iterative technique, Model-Based Incentive Strategy at the Edge (MISE), to improve the overall performance. We use simulations to validate the performance of both, snapshot optimization and MISE, and compare them to a baseline greedy heuristic. MISE strikes a balance between cost-efficiency and worker fatigue (represented in the temporal constraint on the worker) that ensures system availability while retaining workers.

This paper is structured as follows: Section II sheds light on REC and the virtual kiosk model. In Section III, we formulate the optimization problem and quantify the impact of incentives on the availability of the system. In Section IV, we describe MISE as a solution to the formulation. In Section V, we present results and simulations that demonstrate the efficacy of our approach, comparing MISE to both numerical optimizers and cost-effective greedy methods. Finally, in Section VI we conclude.

II. Reward Edge Computing & Virtual Kiosk Model

In the evolving landscape of edge computing, REC systems represent a novel approach utilizing user-owned infrastructure at the extreme edge. Unlike traditional models, RECs incentivize local workers in the immediate vicinity of end-customers through a service provider acting as an orchestrator. This incentive-based approach not only enhances the responsiveness and efficiency of service delivery but also optimizes the use of distributed resources. The fundamental conceptt and operation of an REC these systems were initially outlined in a previous work [2].

A. Reward Edge Computing

REC operates on a semi-distributed model at the extreme edge and is comprised of three main entities:

- **End-users:** Individuals or entities seeking specific services.
- **Worker devices:** User-owned devices that are rented out to provide services in exchange for rewards.
- **Service Providers (Orchestrators):** Manage the deployment of services on worker devices while balancing profitability and reward payouts.

This structure enables service providers to provide services without owning the infrastructure by instead utilizing resources from nearby worker devices.

In the operational framework of REC, the orchestrator's role is pivotal in maintaining the service availability and effectiveness of the services provided. In this work, we focus on a class of RECs that prioritize keeping a service continuously operational (or alive). Orchestrators in such systems actively manage worker devices to ensure a balance between service availability and the incentivization of workers who operate according to IVQ (or an equivalent discipline). The IVQ model, particularly its long-term operational efficiency, was previously explored, but here we extend this to a real-time or online model to capture dynamic aspects discussed in [2]. We upgrade the worker slot model into what we term the "virtual kiosk model," enhancing the flexibility and scalability of service deployment across varied locations.

B. Virtual Kiosk Model

The Virtual Kiosk Model (VKM) is a way to model REC systems that redefines how services are segmented and managed that builds upon the model in [2]. In VKM, service is divided into parallel virtual kiosks, each representing a deployable thread of service that operates autonomously. Workers are recruited by the service provider and arrive at these kiosks according to some arrival rate λ_W. We assume that the total arrival rate is evenly distributed among the K kiosks, so that each kiosk effectively observes λ_W/K arrivals, however it is possible to dynamically adjust such a distribution. Workers occupy these kiosks, engage in service activities governed by the Incentive Vacation Queueing (IVQ) model, and eventually leave after their designated churn time.

To ensure the service remains operational (or "alive"), at least $K_{\min}$ kiosks must be simultaneously occupied. Each kiosk experiences its own sequence of worker arrivals and departures, impacting the overall availability of the service. The uptime of the service is guaranteed only when the number of actively occupied kiosks meets or exceeds $K_{\min}$. This requirement captures the link between availability and the stochastic nature of worker sequences observed by each kiosk. Figure 1 illustrates the the VKM model.

Each kiosk within the VKM operates autonomously, managing its unique sequence of worker operations. The $(w,k)^{\text{th}}$ worker at each kiosk follows the IVQ discipline. Vacation queueing is characterized by alternating periods of service and vacation [12]. In IVQ, the duration of the vacation is determined

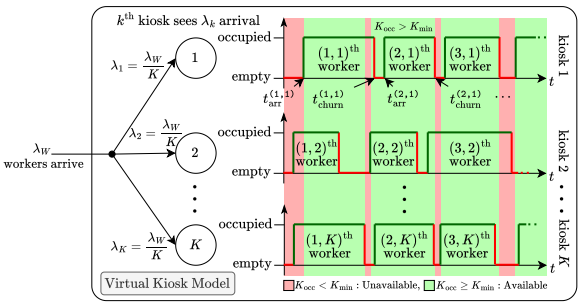

Fig. 1. Virtual Kiosk Model: workers arrive to kiosks in which they provide service to the end-user on behalf of the service provider. The availability of the service is based on the availability of at least $K_{\min}$ workers.

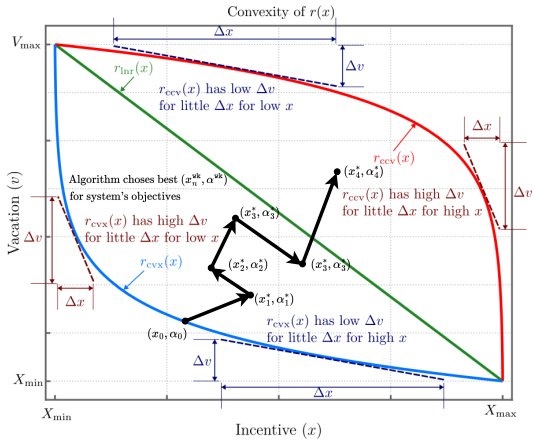

Fig. 2. In IVQ, α affects the pricing of vacation durations, with low α creating a convex V-x curve—less costly short vacation changes at low x and more costly at high x. High α inversely prices small vacation changes—more costly at low x and less so at high x.

by the Incentive-Vacation Function (IVF), $r(x, \alpha)$ [9]:

$$V = r(x, \alpha) = \begin{cases} (1+\alpha) r_{\text{lnr}}(x) - \alpha r_{\text{cvx}}(x) & -1 \le \alpha \le 0 \\ (1-\alpha) r_{\text{lnr}}(x) + \alpha r_{\text{ccv}}(x) & 0 < \alpha \le 1 \end{cases} \tag{1}$$

where the job's incentive x spans $[x_{\min}, x_{\max}]$, and V ranges from $[V_{\min}, V_{\max}]$. The convex and concave IVF, defined as:

$$r_{\text{cvx}}(x) = \frac{V_{\max} V_{\min}(x_{\min} - x_{\max})}{(V_{\min} - V_{\max})x - x_{\max} V_{\min} + V_{\max} x_{\min}}, \tag{2}$$

$$r_{\text{ccv}}(x) = \frac{V_{\max}^2 (x - x_{\max}) + V_{\min}^2 (x_{\min} - x)}{V_{\max}(x - x_{\max}) + V_{\min}(x_{\min} - x)}, \tag{3}$$

adjust the vacation duration based on the incentive amount, impacting the kiosk's operational efficiency and response to workload variations. Figure 2 illustrates the rational family of IVF functions developed in [9], and shows how dynamic incentivization could impact it.

This configuration ensures dynamic responsiveness to the fluctuating worker availability and service demands, maintaining service continuity by requiring at least $K_{\min}$ kiosks to be continuously occupied.

Upon arrival at the k^{th} kiosk, each w^{th} worker is characterized by a set of operational parameters that define their tenure. They arrive at a specific time $t_{\text{arr}}^{\text{wk}}$. Workers set a target profit $X_{\text{target}}^{\text{wk}}$, anticipating churn once this target is achieved to reallocate resources to other endeavors. The duration until churn, $d_{\text{churn}}^{\text{wk}}(x_n^{\text{wk}}, \alpha_n^{\text{wk}})$, is estimated using the formula:

$$d_{\text{churn}}^{\text{wk}} = \frac{X_{\text{target}}^{\text{wk}}}{x_n^{\text{wk}}} \left(S_n^{\text{wk}} + (1 + \alpha_n^{\text{wk}}) \frac{V_{\min} + V_{\max}}{2} - \alpha_n^{\text{wk}} \frac{V_{\min} V_{\max} \log\left(\frac{V_{\max}}{V_{\min}}\right)}{V_{\max} - V_{\min}} \right) \tag{4}$$

and the corresponding churn time $t_{\text{churn}}^{\text{wk}} = t_{\text{arr}}^{\text{wk}} + d_{\text{churn}}^{\text{wk}}$. It is crucial that $d_{\max}^{\text{wk}} > d_{\text{churn}}^{\text{wk}}$, ensuring that the worker's available time exceeds their expected churn duration. The churn rate, $\mu_k(x, \alpha) = 1/d_{\text{churn}}^{\text{wk}}(x, \alpha)$, dynamically adjusts based on the incentives and conditions $(x, \alpha)(t)$. Each worker's service cycle, $\widetilde{S}_n^{\text{wk}}$, includes a service period S_n^{wk} and a subsequent vacation period V_n^{wk}, determined by the incentives x_n^{wk} and the parameter α_n^{wk}, showcasing the nuanced interaction between operational constraints and incentive structures within the kiosk model.

We define an availability flag for the k^{th} kiosk, $A_k(t)$, if it is occupied by a worker where

$$A_k(t) = \begin{cases} 1 & \text{kiosk } k \text{ is occupied by a worker,} \\ 0 & \text{kiosk } k \text{ is empty.} \end{cases}, \tag{5}$$

The total number of occupied kiosks, $K_{\text{occ}}(t)$ can then be obtained

$$K_{\text{occ}}(t) = \sum_{k=1}^{K} A_k(t) \tag{6}$$

which allows us to express the server availability, $A(t)$, as

$$A(t) = \begin{cases} 1 & K_{\text{occ}}(t) \ge K_{\min} \\ 0 & K_{\text{occ}}(t) < K_{\min} \end{cases} \tag{7}$$

In the following section, we formulate the operational dynamics and constraints of VKM as an optimization problem. We then develop a sub-optimal heuristic algorithm to manage worker allocation and resource distribution effectively. Additionally, we introduce a baseline heuristic for comparative analysis to validate the efficacy of our proposed approach.

III. Problem Formulation: Optimal Incentive

Given the VKM and the estimator $d_{\text{churn}}^{\text{wk}}$ that estimates the time the w^{th} worker would spend in the k^{th} slot given the choice of $\{x_n^{\text{wk}}, \alpha_n^{\text{wk}}\}$, it is natural that the server provider would seek to maximize their service's availability while minimizing cost. However, such unconstrained optimization would inevitably lead to the exploitation of worker devices as the natural policy would be to give the minimum incentive and the worst α such that the workers would take as much time as possible to achieve their $X_{\text{target}}^{\text{wk}}$. To constrain this sort of behaviour, a constraint on the worker's amount of work is necessary. To that end, we requires workers to declare a maximum duration $d_{\max}^{\text{wk}}$ they can stay, necessitating departure by time $t_{\max}^{\text{wk}} = t_{\text{arrival}}^{\text{wk}} + d_{\max}^{\text{wk}}$. Having such a temporal constraint, or *deadline* forces the service provider to provide a fair incentive. With this in mind, we proceed to formulating the optimization problem as follows:

$$\begin{aligned} \min_{\substack{x_n^{\text{wk}} \in \mathbf{x}_n \\ \alpha_n^{\text{wk}} \in \boldsymbol{\alpha}_n}} \quad & \sum_{k=1}^{K} x_n^{\text{wk}} \\ \text{subject to} \quad & K_{\text{occ}}(t_n) \ge K_{\min} \\ & t_{\text{churn}}^{\text{wk}} \le t_{\max}^{\text{wk}} \end{aligned} \tag{8}$$

where $\{x_n^{\text{wk}}, \alpha_n^{\text{wk}}\} \in \mathbf{x}_n \times \boldsymbol{\alpha}_n$ represents the decision variables that comprise the decision vectors $\mathbf{x}_n \in [x_{\min}, x_{\max}]^k$ and $\boldsymbol{\alpha}_n \in [-1, 1]^k$ for all kiosks for the n^{th} round.

In this formulation, we minimize the instantaneous incentive for each of the $K_{\text{occ}}(t_n)$ workers occupying the kiosks, while

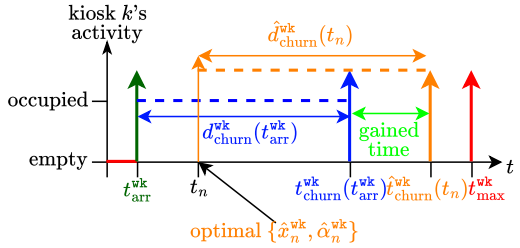

Fig. 3. Pushing the Envelope: Optimal choice of $\{x^{\text{wk}}_n, \alpha^{\text{wk}}_n\}$ seeks to push the worker's churn time from $t^{\text{wk}}_{\text{churn}}$ to a more optimal $\hat{t}^{\text{wk}}_{\text{churn}}$ closer to the worker's maximum time $t^{\text{wk}}_{\text{max}}$; in other words, it seeks to gain more time for the worker in the kiosk.

constraining $K_{\text{occ}}(t_n)$ to be greater than $K_{\min}$, as in the first constraint. The following constraint corresponding prevents the choice of $\{x^{\text{wk}}_n, \alpha^{\text{wk}}_n\}$ that would push the worker to work beyond $t^{\text{wk}}_{\text{max}}$.

The intuition behind the formulation in Eq. 8 is summarized in Figure 3, where the system has an initial estimate of the worker's departure time at $t^{\text{wk}}_{\text{churn}}$ that has been conceived at $t^{\text{wk}}_{\text{arr}}$. By invoking the optimization at the time of some n^{th} round, t_n, we obtain a new estimate of the churn time, $\hat{t}^{\text{wk}}_{\text{churn}}$, based on the optimal choice of $\{x^{\text{wk}}_n, \alpha^{\text{wk}}_n\}$. This estimate is closer to the worker's maximum time $t^{\text{wk}}_{\text{max}}$ than the initial estimate, i.e.,

$$\underbrace{t^{\text{wk}}_{\text{max}} - t^{\text{wk}}_{\text{churn}}}_{\substack{\text{potential availability}\\ \text{lost without optimization}}} < \underbrace{t^{\text{wk}}_{\text{max}} - \hat{t}^{\text{wk}}_{\text{churn}}}_{\substack{\text{potential availability}\\ \text{lost with optimization}}} . \tag{9}$$

A consequence of such an optimal choice is that the service provider experiences a performance gain for each worker in terms of the time they spend in the kiosk, which ultimately translates to service availability.

However, solving such problem using conventional methods, e.g., convex optimization [13], is intractable as the problem is dynamic. In fact, these conventional methods optimize for a snapshot of the system at each round, and will, consequently, always converge to local optima. Hence, in the following section, we propose and devise an online control algorithm that optimizes the choice of $\{x^{\text{wk}}_n, \alpha^{\text{wk}}_n\}$ in a dynamic manner rather than optimizing for a snapshot of the system.

IV. Model-Based Incentive Strategy at the Edge

In this section, we present the Model-Based Incentive Strategy at the Edge (MISE) algorithm, a dynamic approach to selecting $\{x^{\text{wk}}_n, \alpha^{\text{wk}}_n\}$ in REC systems. The MISE algorithm is designed to adjust this choice in real-time to optimize the balance between immediate profitability and long-term sustainability, taking into consideration worker endurance (represented in their maximum work duration), the efficacy of incentives, and the stochastic nature of worker availability. This approach aims to enhance both the reliability of service delivery and worker satisfaction.

We begin by categorizing the workers arriving to the system into two sets prior their arrival to the kiosks: an *elite* set and a *normal* set. The elite set comprises the $K_{\min}$ workers with the highest $t^{\text{wk}}_{\text{max}}$ which are expected to remain the longest. The rest of the workers are then categorized as part of the normal set. The dynamic incentive adjustment process is described as follows:

a) Elite Set Management: For each worker in the elite set, we estimate the time slot $t^{\text{wk}*}_{\text{churn}}$ for which they are likely to stay active in the system. This is calculated using the formula:

$$t^{\text{wk}*}_{\text{churn}} = \frac{X_{\text{target}}}{\bar{x}_{1:n-1}} \delta t \tag{10}$$

where $\bar{x}_{1:n-1}$ represents the average incentive the worker has received since $t^{\text{wk}}_{\text{arr}}$ until the $(n-1)^{\text{th}}$ round t_{n-1}. We assume that the service cycle duration (the duration of a single round) is δt corresponding to the decision horizon of MISE. We then proceed to define a factor, $x^{\text{wk}}_{\text{factor}}$, as:

$$x^{\text{wk}}_{\text{factor}} = \frac{t^{\text{wk}*}_{\text{churn}} - t^{\text{wk}}_{\text{max}}}{t^{\text{wk}}_{\text{max}}} \tag{11}$$

where $t^{\text{wk}}_{\text{max}}$ is the maximum time the worker is available. This factor adjusts the incentive to align the estimated time slot $t^{\text{wk}*}_{\text{churn}}$ with $t^{\text{wk}}_{\text{max}}$, ensuring that the worker does not leave prematurely or stay beyond their maximum available time. The incentive is then updated using:

$$\begin{aligned} x^{\text{wk}}_{n+1} &\leftarrow x^{\text{wk}}_n + \varepsilon_x x^{\text{wk}}_{\text{factor}} \\ x^{\text{wk}}_{n+1} &\leftarrow \text{clip}(x^{\text{wk}}_n, x_{\min}, x_{\max}) \end{aligned} \tag{12}$$

where ε_x is the incentive update step size.

We then proceed to present the resource availability metric ϕ, defined as:

$$\phi(t) = \frac{K_{\min}}{K_{\text{occ}}(t)} \tag{13}$$

This ratio indicates the sufficiency of resources, where a higher ϕ suggests a scarcity of workers, and a low ϕ represents abundance of workers. For the elite set, we will update α^{wk}_n as the following:

$$\begin{aligned} \alpha^{\text{wk}}_{n+1} &\leftarrow \alpha^{\text{wk}}_n - \varepsilon_\alpha \phi \\ \alpha^{\text{wk}}_{n+1} &\leftarrow \text{clip}(\alpha^{\text{wk}}_n, -1, 1) \end{aligned} \tag{14}$$

where ε_α is the α update step size.

The intuition behind this strategy is to prioritize the elite set of workers by varying the parameter α^{wk}_n. If there are insufficient workers, α^{wk}_n is significantly increased for the elite group to encourage them to stay longer, thus ensuring service continuity. Conversely, when there is an abundance of workers, α^{wk}_n is increased less substantially; if an elite worker departs, there are ample others to maintain uninterrupted service.

b) Normal Set Management: For workers in the normal set, we adjust the incentives and parameters differently as follows:

$$\begin{aligned} \alpha^{\text{wk}}_{n+1} &\leftarrow \alpha^{\text{wk}}_n - \varepsilon_\alpha (1-\phi) \\ \alpha^{\text{wk}}_{n+1} &\leftarrow \text{clip}(\alpha^{\text{wk}}_n, -1, 1) \\ x^{\text{wk}}_{n+1} &\leftarrow x^{\text{wk}}_n - \varepsilon_x (1-\phi) \\ x^{\text{wk}}_{n+1} &\leftarrow \text{clip}(x^{\text{wk}}_n, x_{\min}, x_{\max}) \end{aligned} \tag{15}$$

For the normal group of workers, the strategy is to minimize their costs by decreasing both the incentive x^{wk}_n and the parameter α^{wk}_n when worker availability is sufficient to sustain service

availability. The factor ϕ, which reflects resource sufficiency, dictates the rate of these reductions: a gradual decrease occurs when resources are scarce to ensure service continuity, whereas a more substantial reduction is applied when resources are plentiful, optimizing the service costs. In fact, the incentive-vacation function, illustrated in Figure 2, decreasing both x_n^{wk} and α_n^{wk} leads to the least possible operations cost and highest service availability possible.

The MISE algorithm operates continuously in an online fashion, evaluating and adjusting $\{x_n^{\text{wk}}, \alpha_n^{\text{wk}}\}$ at each time step. This process ensures that the amount of incentives attached to the jobs are responsive to changes in worker availability and service demand, dynamically balancing the need to maximize service uptime with the optimization of service costs (in terms of incentives paid to workers). The ultimate goal is to maintain a high level of service availability while ensuring that the workers are adequately compensated and motivated to continue participating in the system. Such a fair approach guarantees that the system satisfies both the end-customers who are seeking its service by ensuring high service availability and the worker devices by giving them opportune profits that can satisfy their corresponding $X_{\text{target}}^{\text{wk}}$.

V. Results

In this section, we simulate the system using the VKM model to investigate the performance of the MISE algorithm in comparison to a constant minimal incentive baseline. We first look at the performance of MISE followed by comparison with snapshot optimization and the baseline. Workers arrive from a worker pool with a rate $\lambda_W = 6$ to the system, and then they are distributed to one of the $K = 20$ slots. For the service to be available, a minimum of $K_{\text{min}} = 5$ workers are required. As for the parameters of the system, $x \in [1, 20]$, $V \in [0.01, 0.05]$ (corresponding to 10% to 50% of the average service cycle in the simulation). We start from initial point $(x_n^{\text{wk}}, \alpha_n^{\text{wk}}) = (10.5, 0)$.

We look at the general distribution of the differences in time and incentives accumulated. The difference in time (in Figure 4) is important for the service provider as it represents the amount of time gained, while the difference between the target incentive and the accumulated incentives (in Figure 5) are important for the worker. In Figure 4, it can be seen that the system seeks to keep the workers in the Elite Set as much as possible to their $t_{\text{max}}^{\text{wk}}$. In contrast, the churn times for other workers are much earlier than their corresponding $t_{\text{max}}^{\text{wk}}$. A similar pattern can be noticed in Figure 5. This distinction between Elites and Normal workers allows the service provider to keep providing a range of incentives slightly higher than the midpoint $(X_{\text{min}} + X_{\text{max}})/2$ while tuning α to valuate the vacation for the elite workers at a fair rate that is consistent with their maximum departure time.

In Figure 6, the number of actively occupied kiosks over time is plotted. The red line represents K_{min}, the service is unavailable when $K_{\text{occ}}(t)$ falls below K_{min}. In Figure 6, it can be seen that the snapshot optimization results in the worst availability as it seeks to minimize the cost with availability as a hard constraint. In contrast, the greedy approach ensures that the system is always available. Compared to both, MISE is superior to the optimization approach in guaranteeing availability, but inferior to the greedy approach. However, in Figure 7, the optimization and greedy approach have a wide cost variation while optimization has a minimal variation in cost. MISE, on the other hand, has median variation in comparison to both.

Figures 8 and 9 demystify the behaviours in Figures 6 and 7. The optimization approach seeks to satisfy as much workers while the greedy seeks to employ them for as long as possible. MISE, on the other hand, seeks to provide median performance between both.

VI. Conclusion

In this study, we have introduced the Virtual Kiosk Model (VKM) and the Model-based Incentive Strategy at the Edge (MISE) to quantify the impact of incentives on the availability of a service deployed on user-owned devices at the extreme edge. By leveraging the idle capacities of user-owned worker devices, the extreme edge can provide service at latencies that are not normally achievable for the traditional enterprise-centric

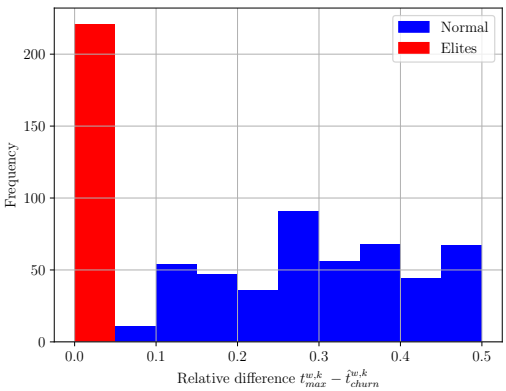

Fig. 4. Ungained Worker Availability Time

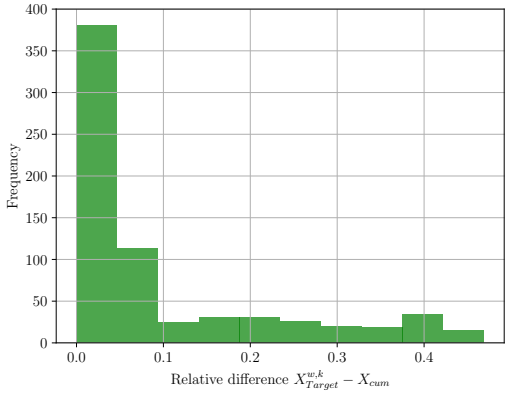

Fig. 5. Unachieved Profits for Workers

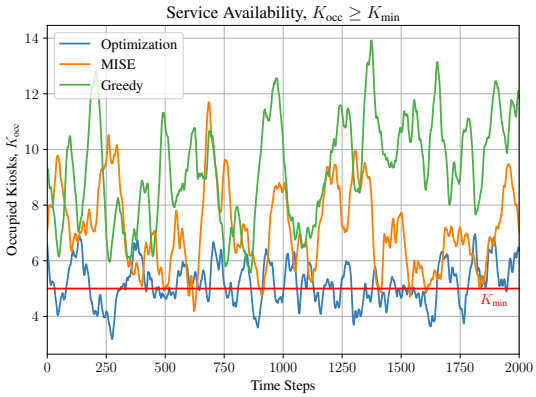

Fig. 6. Service Availability in terms of $K_{\text{occ}}(t)$

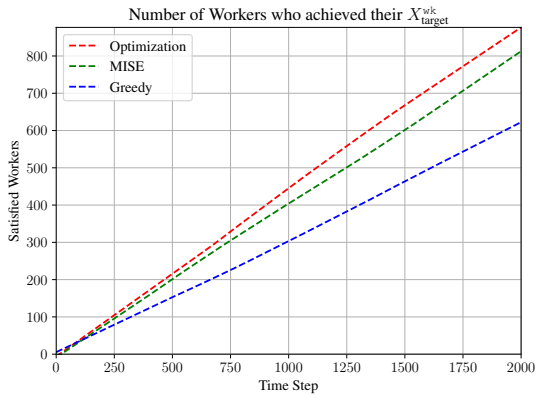

Fig. 8. Number of Satisfied Workers who have achieved their desired $x^{\text{wk}}_{\text{target}}$

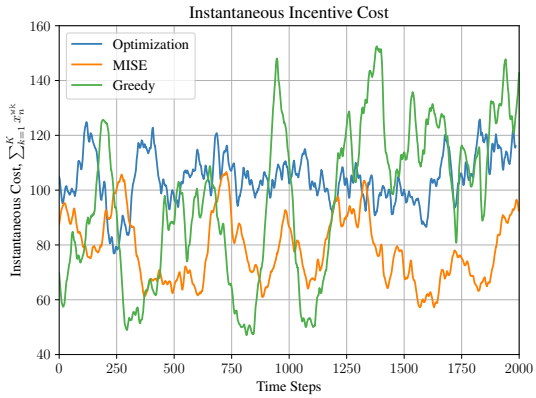

Fig. 7. Instantaneous Incentive Cost incurred by Service Provider

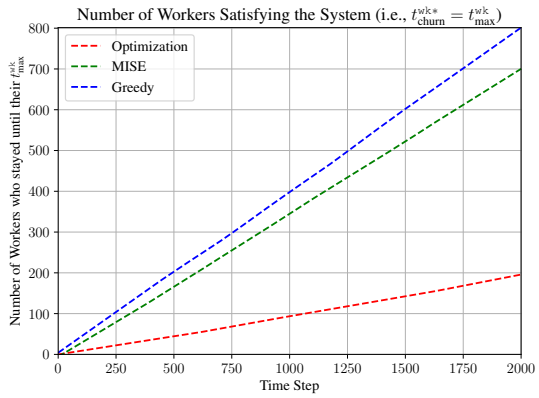

Fig. 9. Number of workers who satisfied the system by achieving their $t^{\text{wk}}_{\text{max}}$

edge computing. Our objective in this work was to study the service availability aspect in a REC setting. However, a trade-off between the service provider and the workers exists. An underpaid yet overworked worker would never return to such a system, which would significantly affect long term service availability due to lack of worker retention for future instances. On the other hand, satisfying workers all the time does not guarantee maximal performance and comes at a significant cost to the service provider. To that end, we developed MISE to address this trade-off dynamically to mitigate the negative effects while ensuring availability and validated its efficacy via simulation.

Acknowledgement

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant ALLRP 549919-20; in part by Distributive, Ltd; and by Qatar University under Grant IRCC-2024-494.

References

[1] Y. Wang, L. Chen *et al.*, "Research on Privacy Preserving Computing Technology in Edge Computing," *2023 International Conference on Networking and Network Applications (NaNA)*, 2023.

[2] S. B. Azmy, N. Zorba, and H. S. Hassanein, "Incentive-Vacation Queueing in Extreme Edge Computing: An Analytical Reward-Based Framework," *IEEE Open Journal of the Communications Society*, vol. 5, 2024.

[3] F. Moebius, T. Pfandzelter, and D. Bermbach, "Are Unikernels Ready for Serverless on the Edge?" *arXiv*, 2024.

[4] A. J. Ferrer, J. M. Marqués, and J. Jorba, "Ad-hoc Edge Cloud: A framework for dynamic creation of Edge computing infrastructures," *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, 2019.

[5] E. Labs, "Exo: Run your own ai cluster at home with everyday devices," https://github.com/exo-explore/exo/tree/main, 2024, accessed: Aug. 8, 2024.

[6] Q. Cai, Y. Zhou *et al.*, "Collaboration of heterogeneous edge computing paradigms: How to fill the gap between theory and practice," *IEEE Wireless Communications*, vol. 31, no. 1, p. 110–117, 2024.

[7] X. Wang, J. Ye *et al.*, "Mean field graph based d2d collaboration and offloading pricing in mobile edge computing," *IEEE/ACM Transactions on Networking*, vol. 32, no. 1, p. 491–505, 2024.

[8] H. Bornholdt, K. Röbert *et al.*, "Measuring the Edge: A Performance Evaluation of Edge Offloading," *2023 IEEE PerCom Workshops*, 2023.

[9] S. B. Azmy, N. Zorba, and H. S. Hassanein, "Incentive-Vacation Queueing for Edge Crowd Computing," *IEEE Internet of Things Journal*, vol. 11, no. 8, 2024.

[10] X. Xu, Q. Cai *et al.*, "An incentive mechanism for crowdsourcing markets with social welfare maximization in cloud-edge computing," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 7, 2021.

[11] D. P. Dias, J. Simão, and L. Veiga, "RATEE - Resource Auction Trading at Edge Environments," *2021 IEEE 20th International Symposium on Network Computing and Applications (NCA)*, 2021.

[12] N. Tian and Z. G. Zhang, *Vacation Queueing Models Theory and Applications*, ser. International Series in Operations Research & Management Science. Springer, 2006.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.