ALLOCATION OF COMPUTING TASKS IN DISTRIBUTED MEC SERVERS CO-POWERED BY RENEWABLE SOURCES AND THE POWER GRID

Davide Cecchinato[*] *Michele Berno*[*] *Flavio Esposito*[†] *Michele Rossi*[*]

[*] Dept. of Information Engineering, University of Padova, Italy
[†] Dept. of Computer Science, Saint Louis University, Saint Louis (MO), US
Email: {cecchin4, michele.berno, rossi}@dei.unipd.it, espositof@slu.edu

ABSTRACT

We consider a Multiaccess Edge Computing (MEC) network where distributed servers have energy harvesting (e.g., solar) and storage (e.g., batteries) capabilities. Energy from a connected *power grid* is also available, in case that harvested from ambient sources is scarce or absent. Network processors are deployed according to a given network topology, across two tiers, and computing tasks are flexibly allocated depending on considerations related to load balancing, energy consumption (for communication and computing) and energy purchases from the power grid. Specifically, an *online* optimization problem, exploiting a predictive control approach, is formulated to minimize the monetary cost incurred in the energy purchases from the power grid, by dispatching the computation jobs to those servers that have enough energy and computation resources. Our proposed framework uses forecasts of exogenous processes, such as the amount of energy harvested and job arrivals, which are estimated on the fly to steer the allocation of computation jobs to the servers.

***Index Terms*—** Renewable Edge Networks, Sustainable Edge Computing, Energy Harvesting.

1. INTRODUCTION

Due to the growth of the communication and computing demands (such as for future vehicular related applications), communication technologies could drain as much as 51% of global electricity by 2030, as reported in [1]. For this reason, and motivated by energy cost saving considerations and environmental concerns, telecommunications operators are considering the deployment of renewable energy sources to supplement conventional energy sources in powering Base Stations (BSs) [2].

In this paper, we formulate an optimization problem aimed at scheduling the computation load, while minimizing the cost due to the energy purchases from the power grid incurred by a telecommunications operator in a Multiaccess Edge Computing (MEC) network, where computing servers have energy harvesting capabilities. This problem is tackled exploiting predictions of harvested energy and computation load via a Model Predictive Control (MPC) approach, obtaining cost savings higher than 50% with respect to myopic strategies.

Related work: In the literature, there exist various works coping with the management of access networks powered by renewable energy sources and the power grid. The authors of [3] consider a scenario with two BSs sharing their energy through a power line and whose aim is to minimize the power purchased from the power grid. A similar scenario with more BSs is investigated in [4, 5], where BSs share their energy to seek self-sustainability and energy independence from the power grid. The objective of minimizing the power obtained from the grid is also pursued in [6]. Here, the authors increase the cell size of those BSs having more harvested energy availability and reduces the cell size of the BSs with low energy availability. In this way, energy-rich BSs would serve more users and spend more energy than energy-poor ones. In our work, we aim at minimizing the *monetary cost* of the purchased energy, which differs from minimizing the amount of energy consumed/purchased. BSs cannot share their energy nor modify their cell size, but they can share the computation tasks that they have to process. In [7], the authors also consider the same objective of minimizing the energy cost, by predicting the energy demand and buying electricity one day ahead. However, they do not consider energy harvesting nor accumulation (e.g., batteries) systems.

This paper is organized as follows. In Section 2 we describe our system model, formulating the optimization problem and solving it in an online fashion through an MPC-based approach. In Section 3 we show some selected results and we quantify the monetary savings that can be achieved with our framework. Finally, in Section 4 some concluding remarks are given.

2. SYSTEM MODEL

The considered network's architecture is composed of Edge Servers (ESs) (set $\mathcal{M}$) equipped with computing and energy harvesting capabilities, see Fig 1. ESs belong to one of the two following layers: tier-1 servers are co-located with BSs

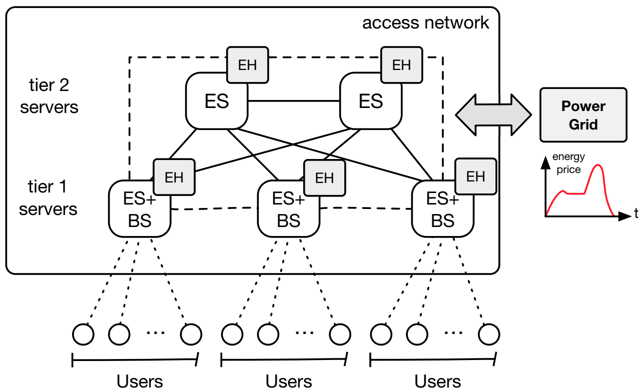

Fig. 1. Renewable edge network architecture. It is composed of two layers of computing enabled nodes, equipped with energy harvesting devices. They can communicate and exchange jobs between each other. Only ESs belonging to tier 1 are connected to BSs and can gather processing requests from users (jobs).

and can receive computation tasks from the co-located BS or from neighboring ESs (located at tiers 1 and 2), while tier-2 nodes are not co-located with BSs and, in turn, can only process jobs received from neighboring ESs (we consider a full mesh topology among first and second tier nodes). The internal architecture of an ES is shown in Fig 2. Edge nodes are composed of several components: a processing unit, an energy buffer that can store energy harvested from the environment or purchased from the grid (with a maximum capacity of $B_i^{\max}$ [kWh], $i \in \mathcal{M}$), a traffic dispatcher that separates the traffic destined to local processing from the one that has to be offloaded to other ESs, and lastly, a communication unit.

Time is slotted, with a slot duration of one hour. At every time slot k, each node $i \in \mathcal{M}$ requires an amount of energy $c_i(k)$ to perform the needed operations (computation and communication). (Throughout the paper energy is measured in Wh.) The nodes' energy sources are the power grid and the energy harvesting system and nodes' batteries can be used either as energy sources or sinks, depending on the time slot. We denote by $d_i(k)$ the energy drained by node i from its battery at time slot k, and by $h_i(k)$ the energy harvested by node i during time slot k. $h_i(k)$ is split into two components: the first one, $h_i^b(k)$, is the portion of harvested energy used for charging the battery; the second one, $h_i^c(k)$, is the portion of harvested energy immediately used by the node to sustain its operations. This is encoded through the following equation,

$$h_i(k) = h_i^b(k) + h_i^c(k). \tag{1}$$

The same reasoning is applied to the energy purchased by node i in time slot k, denoted by $g_i(k)$, which is split into two components: the first one, $g_i^b(k)$, representing the portion of purchased energy that is used for charging the battery; the second one, $g_i^c(k)$, representing the portion of purchased

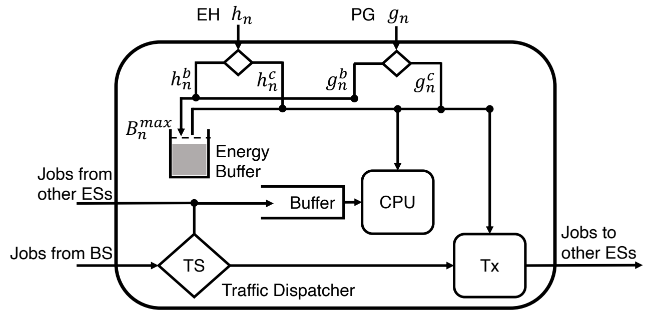

Fig. 2. Edge servers' internal architecture. Each ES is equipped with an energy harvesting device, a computing unit, a traffic dispatcher, and a communication unit.

energy that is immediately used by the node to sustain its operations. It follows that,

$$g_i(k) = g_i^b(k) + g_i^c(k). \tag{2}$$

In the interest of optimizing the network energy and computation resources, ESs can cooperate by offloading (in full or in part) the jobs collected from the BSs. We define $r_{ij}(k)$ (jobs/hour) as the job flow that node $i \in \mathcal{M}$ transfers (offloads) towards node j (one among its neighboring nodes) in time slot k. Note that $r_{ii}(k)$ represents the portion of job flow that is directly collected by ES i from the co-located BS. We model the energy required by node i in time slot k as

$$c_i(k) = \alpha \left(r_{ii} + \sum_{j \in \mathcal{M} \setminus \{i\}} r_{ji} \right) + \beta \left(\sum_{j \in \mathcal{M} \setminus \{i\}} r_{ij} \right). \tag{3}$$

Note that $\alpha > 0$ while $\beta < 0$ (with $|\alpha| > |\beta|$).

The amount of energy $c_i(k)$ can be gathered from the power grid, the harvesting system and/or the node's battery. In particolar, it must hold

$$c_i(k) = g_i^c(k) + h_i^c(k) + d_i(k). \tag{4}$$

The battery evolution is described as

$$b_i(k) = \eta_i \left(b_i(k-1) - d_i(k) \right) + \mu_i \left(h_i^b(k) + g_i^b(k) \right), \tag{5}$$

where $\eta_i \in (0, 1]$ is a parameter accounting for the self-discharging behavior of the battery, whereas $\mu_i \in (0, 1]$ accounts for the losses in the charging process.

There are other constraints in addition to (1), (2), (3), (4), and (5), as detailed next.

Non-negativity: the variables (and, actually, all the considered signals) are non-negative

$$g_i(k) \geq 0, \ \forall i, \ \forall k. \tag{6}$$

Maximum battery capacity: the battery charge level must be smaller than or equal to the battery capacity

$$b_i(k) \leq B_i^{\max}, \ \forall i, \ \forall k. \tag{7}$$

Maximum battery drained: the energy drained from the battery must be smaller than or equal to the battery charge level

$$d_i(k) \leq b_i(k-1), \ \forall i, \ \forall k, \tag{8}$$

where with $b_i(0)$ we mean the initial battery level of node i.

Maximum harvested energy: the actual amount of harvested energy cannot exceed the maximum amount of energy that could be harvested from the environment at a certain time slot. In formulas, this means that

$$h_i^c(k) + h_i^b(k) \leq H_i^{\max}(k), \ \forall i, \ \forall k. \tag{9}$$

Flow conservation: at every time slot, the number of jobs exiting a node cannot exceed the number of jobs entering the same node

$$\sum_{j \in \mathcal{M} \setminus \{i\}} (r_{ij}(k) - r_{ji}(k)) \leq r_{ii}(k), \ \forall i, \ \forall k. \tag{10}$$

Additional flow constraints: nodes have a maximum flow processing constraint per time slot

$$r_{ii} + \sum_{j \in \mathcal{M} \setminus \{i\}} (r_{ji}(k) - r_{ij}(k)) \leq R_i^{\max}(k), \ \forall i, \ \forall k. \tag{11}$$

2.1. Optimization problem

We formulate an optimization problem with the objective of minimizing the monetary cost of the purchased energy, while meeting the above constraints,

$$\underset{g_i(k),\, i \in \mathcal{M},\, 0 \leq k \leq T}{\text{minimize}} \quad \sum_{r=1}^{T} p(r) \sum_{i \in \mathcal{M}} g_i(k) \tag{12}$$

subject to: Eqs. (1), (2), (3), (4), (5), (6), (7), (8), (9), (10), (11).

2.2. Model Predictive Control approach

Solving problem (12) entails a complete knowledge of harvested energy arrivals, job flows, and energy price across all (T) time slots. Since in a real scenario such knowledge is impossible to achieve (in fact, we only have knowledge of the processes in the current, k, and past time slots), we adopt a more practical MPC approach to solve problem (12) in an online fashion [8]. Specifically, at every time slot k, we compute predicted values of harvested energy arrivals, job flows and energy price for the next $W < T$ time slots, and we solve (12) for these W time slots using such predictions (see Section 2.3 for additional details on forecasting). The solution of this problem specifies the energy purchases and the computation flow dispatchment for slots $k, k+1, \dots, k+W-1$. Out of these, according to the *receding horizon principle*, only the actions for the current slot k are implemented, whereas those associated with future time slots, $k+1, \dots, k+W-1$, are discarded. At time $k+1$, the predictions are updated and this procedure is iterated.

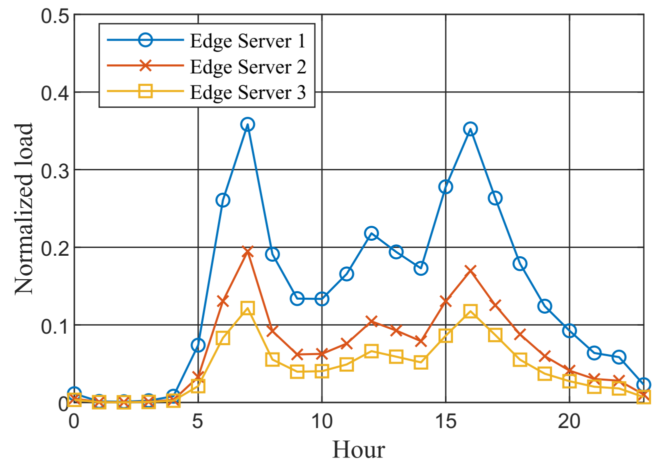

Fig. 3. The three load profiles used in the simulations. Notice that the lines exhibit spikes approximately in correspondence to the hours of the day when most people commute.

2.3. Prediction of exogenous processes

Three exogenous processes are considered in our framework: the energy price $p(k)$, the harvested energy arrivals $h_i(k)$ and the loads $r_{ii}(k)$ (i.e., the job flows to be processed). In this section, we describe how the traces used in the simulations were collected and how their predictions were performed.

Energy price: hourly electric energy prices were gathered from the US National Grid database [9] (we collected the prices for year 2018). Usually, energy prices are available one day ahead ("one day ahead Market"), so their prediction is not needed if we consider a time horizon within 24 hours.

Harvested energy: we used the SolarStat tool [10], that comes with real solar energy traces collected over 20 years across the city of Los Angeles. To generate harvested energy predictions we adopted Long Shot-Term Memory (LSTM) neural networks [11]. Our LSTM-based predictor has been trained to directly output the forecasts for the required number of future time slots W. We trained an LSTM network with one hidden layer consisting of 40 neurons, for 80 epochs over 4 years of harvested energy measurements, see also [12].

Job flows: job flows $r_{ii}(k)$ are obtained considering a vehicular edge computing scenario. We assume that the job flow arriving at server i is directly proportional to the vehicular traffic present in the areas served by the BSs connected to that server. To obtain numerical time series, we used the mobility simulator SUMO [13] to track the number of vehicles in different areas of the city of Cologne during an entire day. Averaging and normalizing the time series obtained in 30 days for 3 different areas of the city (each of 200×200 square meters), we obtained the load profiles for the three tier-1 servers (see Fig. 3). (Normalization is performed by dividing by the maximum number of vehicles that were simultaneously present in

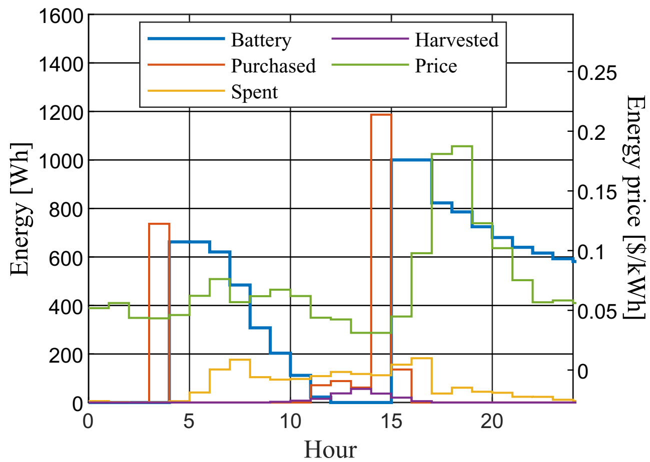

Fig. 4. Telemetry of tier-1 server 1 energy processes plus energy price process across 24 hours. Note that purchases are made when the energy price exhibits local minima.

Name	Value
α	0.108 Wh/(job/h)
β	-0.036 Wh/(job/h)
η_i	0.9999
μ_i	0.900

Name	Value
T_{sim}	360 h
$B_i^{\max}$	1.00 kWh
$b_i(0)$	0.00 Wh
$R_i^{\max}$	3330 job/h

Table 1. Simulation parameters. Parameters depending on i apply to all nodes $i \in \mathcal{M}$.

the areas.) Such load profiles are used as job flow predictions in the considered MPC optimization approach.

3. RESULTS

Next, we present some selected results that have been obtained simulating the system described in section 2 using the parameters reported in Table 3. In particular, we compare the performance obtained through three different approaches: the globally optimal (offline) approach (Eq. (12) with full information across all time slots), the genie-aided MPC (Genie-MPC) and the predictor-based MPC (Pred-MPC). The globally optimal solver has knowledge of the exogenous processes over the whole simulated interval T_{sim}; the genie-aided MPC method has knowledge about future arrivals over W time slots; the predictor-based approach uses load profiles and LSTM-based energy forecasts, as discussed in section 2.3.

Energy purchase decisions: Fig 4 shows the energy price and the state evolution of server 1 (in tier 1) obtained using Pred-MPC with $W = 24$ hours. Notably, the ES purchases energy from the power grid in correspondence of energy price's local minima, thus reducing the incurred cost.

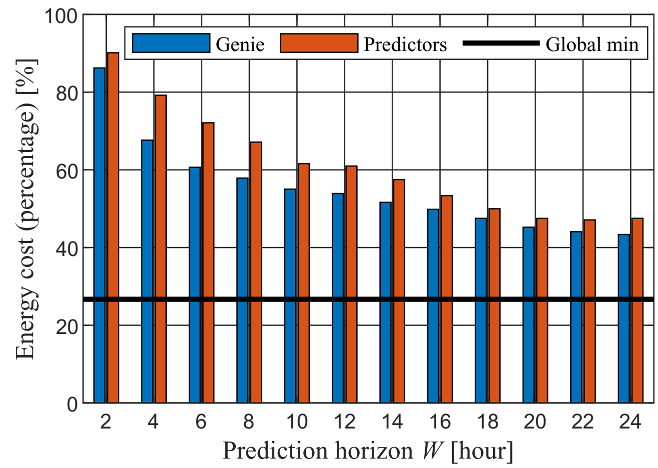

Fig. 5. Comparison in terms of energy cost between Genie-MPC and Pred-MPC while varying W. Results are reported as percentage with respect to the myopic approach ($W = 1$). The black line indicates the globally optimal energy cost.

Energy cost savings: Fig. 5 shows the energy cost achieved by Genie-MPC and Pred-MPC while varying the length of the prediction horizon W. Results are expressed as a percentage with respect to the energy cost incurred with $W = 1$ (corresponding to a myopic allocation of jobs). Note that, when $W = 1$, the system makes decision based only on the job and energy arrivals for the current time slot, without exploiting predictions and without considering energy price fluctuations (and, indeed, Genie-MPC and Pred-MPC coincide for $W = 1$). The black horizontal line represents the globally optimal solution's cost: perfect knowledge of all processes at all time slots and its exploitation in the optimization ideally allows to save up to 70% of the overall energy cost with respect to a myopic allocation strategy. In a more realistic scenario, setting $W = 4$ hours, allows reducing the energy expenses by 20%. With $W = 18$ hours, the energy cost is halved using Pred-MPC. Finally, we remark that increasing W the performances of Pred-MPC are close to that of Genie-MPC.

4. CONCLUSIONS

We have considered a MEC network where servers are distributed across two tiers and have energy harvesting capabilities. Energy from a connected power grid is also available, in case that harvested from ambient sources is scarce or absent. For this scenario, we advocated the use of predictive control to steer, at runtime, the allocation of computation jobs to the servers, and the amount of energy that the MEC network purchases from the power grid. We showed that predictive control provides considerable savings on the overall energy cost incurred (up to 50%), even when imperfect estimates are available for future energy and traffic arrivals.

5. REFERENCES

[1] A. Andrae and T. Edler, "On global electricity usage of communication technology: trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.

[2] "Building Zero Emissions Radio Access Networks: Nokia White Paper," Tech. Rep., Nokia, 2016.

[3] Y. Chia, S. Sun, and R. Zhang, "Energy cooperation in cellular networks with renewable powered base stations," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6996–7010, Dec 2014.

[4] A. F. Gambin and M. Rossi, "Energy cooperation for sustainable base station deployments: principles and algorithms," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–7.

[5] A. F. Gambin, M. Scalabrin, and M. Rossi, "Online Power Management Strategies for Energy Harvesting Mobile Networks," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 721–738, 2019.

[6] T. Han and N. Ansari, "On optimizing green energy utilization for cellular networks with hybrid energy supplies," *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, pp. 3872–3882, 2013.

[7] J. Xu, L. Duan, and R. Zhang, "Energy group buying with loading sharing for green cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 786–799, April 2016.

[8] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*, Nob Hill Publishing, 2015.

[9] US National Grid, "Electric Supply Charge," Online, Accessed on 2019-10-11.

[10] M. Miozzo, D. Zordan, P. Dini, and M. Rossi, "Solarstat: Modeling photovoltaic sources through stochastic markov processes," in *2014 IEEE International Energy Conference (ENERGYCON)*, May 2014, pp. 688–695.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT press, 2016.

[12] T. Dlamini, A. F. Gambin, D. Munaretto, and M. Rossi, "Online Supervisory Control and Resource Management for Energy Harvesting BS Sites Empowered with Computation Capabilities," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.

[13] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, pp. 128–138, December 2012.