Journal of Economy and Technology 1 (2023) 208–221

Contents lists available at ScienceDirect

Journal of Economy and Technology

journal homepage: www.keaipublishing.com/en/journals/journal-of-economy-and-technology/

Next generation edge computing: A roadmap to net zero emissions

Raghubir Singh[a], Sukhpal Singh Gill[b,*]

[a] Department of Computer Science, University of Bath, United Kingdom
[b] School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom

ARTICLE INFO

Keywords:
Edge Computing
Net zero emissions
Task offloading
Computational complexity
Wireless Local Area Networks
Energy-efficiency

ABSTRACT

Multi-access Edge Computing (MEC, formerly, Mobile Edge Computing) offers users of Mobile Devices (MDs) such as smartphones advantages in computational task completion times by off-loading to superior but geographically close computing resources. There is a need to make these services more sustainable (energy-efficient) as an initial step towards net zero emissions. We have modelled a zero-latency edge node-MD connection as a single base station linked to a devoted server with the MD transmitting and receiving data via a Wireless Local Area Network (WLAN) with, wherever possible, identifiable real-world power ratings. In this paper, we demonstrate that the total energy usage by the MD and edge node hardware can be markedly less than that of the MD alone in 3 G, 4 G and 5 G WLAN networks. The energy savings, computed as percentages of MD-alone energy usage, are independent of data file size, are greater with computationally more complex tasks, and have a higher MD CPU workload but moderate server CPU workloads. Energy savings are highly dependent on the precise configuration of the base station with server type (peak power rating, the number of CPU cores and the maximum number of jobs processed simultaneously). A 5 G base station has the highest power consumption, but this is offset by much faster WLAN speeds, which can result in energy savings in excess of 90% compared with MD computation alone. We discuss the implications of these results for reducing global electricity use and achieving carbon neutrality to contribute towards net zero emissions.

1. Introduction

Communication technology is predicted in some scenarios to account for 50% of global electricity demand and 23% of globally released greenhouse gas emissions by 2030 Andrae and Edler (2015), which needs to be controlled to achieve carbon neutrality to contribute towards net zero emissions Cao et al. (2022). While important drivers for this trend of global connectivity include the Internet of Things (IoT) and Cloud Computing, a significant contributor is the increasing installation of devices controllable by smartphones Singh et al. (2023). The evolution of the smartphone has required increasing power ratings for applications such as location sensing Zhuang et al. (2010). Subsequently, tablet computers have become an increasingly popular form of a computational mobile device (MD) with many applications in higher education (Gill et al., 2024b; Ramaraj, 2016. Users of MDs, therefore, are open to innovative developments in energy efficiency Seto et al. (2021), which put individual choices in a framework of economic costs, lifestyle choices and environmental awareness, especially in terms of carbon emissions (Walia et al., 2023; Gill et al., 2024a. So, there

* Correspondence to: School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, United Kingdom.

E-mail address: s.s.gill@qmul.ac.uk (S.S. Gill).

https://doi.org/10.1016/j.ject.2023.12.001
Received 27 September 2023; Received in revised form 6 December 2023; Accepted 7 December 2023
Available online 11 December 2023

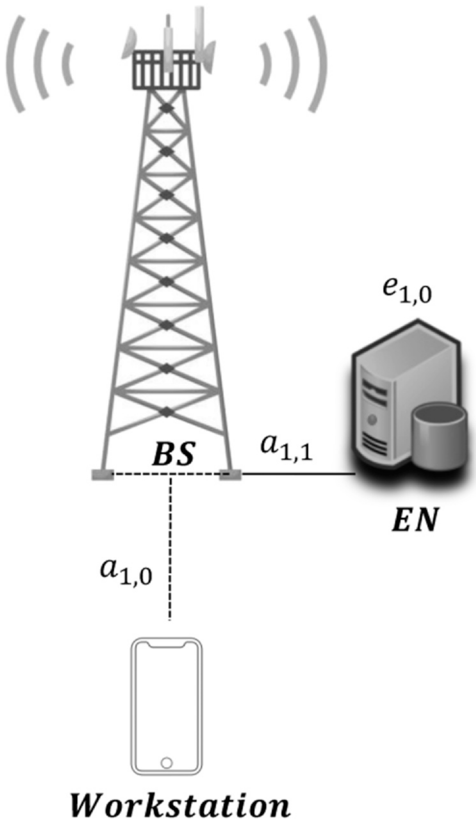

Fig. 1. A typical MEC network structure.

is a need to make mobile edge computing services more sustainable (energy-efficient and cost-effective) Kurniawan et al. (2023) and environmentally friendly as an initial step towards net zero emissions Senthilkumar et al. (2023).

Since 2014, users of MDs have been the subjects of a paradigm shift in computing originally known as Mobile Edge Computing and subsequently as Multi-access Edge Computing Sabella et al. (2020). Edge Computing, in its broader definition, offers human users and autonomic (IoT) devices access to superior computing capabilities housed "locally" in edge node servers (and ancillary hardware) as opposed to being concentrated in geographically distant cloud data centres to minimise latency delays (Sabella et al., 2020, 2016; Safavat et al., 2020; Malandrino et al., 2016; Liu et al., 2018; Varghese et al., 2016. Computational offloading from MDs has been the focus of multiple studies to define the advantages of using edge nodes to shorten computation time and access expanded computing horizons (Liu et al., 2017; Lin et al., 2019; Jiang et al., 2019; Zhang and Zhao, 2020. Not only can computing times be greatly shortened by offloading to edge nodes, energy use by MDs is also reduced (Varghese et al., 2016; Sonmez et al., 2017; Du et al., 2023; Singh et al., 2019; Iftikhar et al., 2022; Singh et al., 2020a, 2021. Whether or not the total energy usage by offloading is less than that of an isolated MD, however, requires a more complex analysis of edge node-MD synergy in an edge computing network with its hierarchy of power-rated devices (Fig. 1). In this paper, we consider a typical MEC network structure consisting of a mobile device (MD) and a MEC server situated at a base station (BS).

The rest of this paper is organised as follows. Section 2 discusses the related work. Section 3 presents the contributions of this work. Section 4 describes the quantitative methodology used. Section 5presents the results of numerical simulations with different base station-server combinations, WLAN speeds, task complexities and task data set sizes. Finally, Section 6 considers the possible horizons for the economies of scale offered by MEC networks for reducing energy use in offloading from MDs and outlines possible future work in this area.

2. Related work and motivation

Previous works on energy usage in offloading tasks to MEC resources can be differentiated into two main focuses: firstly, only on energy use by MDs during offloading and, secondly, studies integrating energy use in both MDs and edge node resources Walia et al. (2023). Among the first group, several studies have shown that offloading can be advantageous for the users of MDs in reducing energy use (and potentially prolonging battery use-time before recharging) with both edge computing networks (Liu et al., 2017; Singh et al., 2021; Mao et al., 2016 and cloud computing (Guo et al., 2019; Wen et al., 2012. The most detailed analysis of the total energy required in offloading, i.e., taking into account the energy used by the edge node as well as the MD, demonstrated large (60–70%) reductions in energy use with relatively small offloaded tasks, but larger tasks could not maintain this energy efficiency

Mahenge et al. (2022). The modelled system, however, involved multiple base stations and servers, and the degraded performance with larger data sets most likely resulted from multiple data transfers in a complex MEC network (which could also merge with cloud resources); data migrations inevitably have energy costs, and this could negate any overall energy savings by offloading to the collaborative MEC network Nandhakumar et al. (2023). Considering only latency-critical applications, a joint offloading and resource allocation strategy for multiple users found only small energy savings (4.5–5.5%) by offloading to a multi-server MEC network Cheng et al. (2018).

3. Our contributions

Differentiating our approach from previous studies, to structure energy use when an MD offloads tasks to an edge node, we have used (wherever possible) data and estimates from real-world hardware components such as base stations and servers (in the edge node) and smartphone power ratings to include the various energy-requiring events implicit in offloading tasks. We have focused the analysis on the MD-edge node collaboration without reliance on any edge-cloud task data communication (Fig. 1). Furthermore, we have assumed that round trip times from MD to edge nodes are insignificantly short relative to the latency demands of the MD application operations, which is a key factor in the adoption of all edge computing modes (Clinch et al., 2012; Hu et al., 2016. Furthermore, we focus on an implementation mode for edge computing resources which links MD users to service providers by Service Level Agreements (SLA) to prioritise subscribers and avoid queuing delays (Fantacci and Picano, 2020; Lin et al., 2022. Finally, we model the MD-edge node relationship as single-tier events with a dedicated multiple core-server to maximise computational flexibility without data transfer times between multiple servers in a collaborative cluster (Wang et al., 2020; Mahenge et al., 2019; Yu and Langar, 2019.

With this clearly defined system architecture, we demonstrate important energy savings possible by offloading from resource-constrained MDs and, in particular:

- Major energy savings compared with energy usage by MDs can be achieved by offloading tasks to edge node servers with 3 G, 4 G or 5 G base stations with a wide variation in Wireless Local Area Network (WLAN) speeds and these savings are maximised with 5 G base stations and 5 G-comparable WLAN speeds.
- These energy savings are task size-independent but highly dependent on the characteristics of the edge node server, in particular, the server peak power, the number of CPU cores and the maximum number of jobs processed simultaneously by the server. Energy efficiency was critically linked to the "economies of scale" possible in edge nodes despite their hardware being power-hungry compared to individual MDs.
- From the MD side, the increasing computational complexity of the offloaded task and higher CPU workload in the MD increased the energy savings made possible by offloading to a computationally superior MEC server.

4. Research methodology

While computational tasks on a mobile device (MD, a smartphone, tablet computer or laptop) require only local energy usage, offloading to a MEC server involves sequential energy-requiring steps: .

1) Data transmission to a base station connected to a server,
2) Computation on a server,
3) Data return to the user,
4) "Background" energy-requiring steps in the base station during the whole time required for task completion

Comparing energy usage in local and offloading modes suggests that achieving any reduction from the local baseline would require economies of scale attainable in MEC hardware capable of simultaneously handling multiple (hundreds or thousands of) tasks, i.e., achieving low energy requirements per individual task Gill et al. (2024a). The symbols used in this section are defined in Table 1.

The two major power-rated items of hardware in a MEC system are the base station hardware components and a connected server. A recent estimate of the power consumption by a 5 G base station emphasised the increase over 4 G network equipment Dongxu (2020): 11577 W (5 G), 6877 W (4 G) and 4808 W (2/3 G). This provides a platform for energy usage computations for base station activities distinct from those of the MEC server.

Table 1
Notations and its definition.

Symbol	Definition
PR	Power rating of a MEC base station
MPR	Minimum power rating of a MEC base station
P_{transmit}	MD power when transmitting data
P_{receive}	MD power when receiving data
P_{compute}	MD power when computing
P_{idle}	MD idling power

Table 2
Server Power per Task (Individual Computational Job); data modified from Ji et al. (2021).

Server Type	Server Power Idle (W)	Server Power Peak (W)	CPU cores	Maximum Jobs	Server Power peak per task (W)
A	100	200	8	2000	0.1000
B	110	300	96	24000	0.0125
C	200	300	32	8000	0.0375
D	200	500	16	4000	0.1250
E	430	1000	64	16000	0.0625
F	1590	2490	24	6000	0.4150

Higher usage of the base station would, however, be expected to require higher energy usage. While no analysis of base station power versus usage statistics is available, considerable attention has been paid to telco base stations (Arnold et al., 2010; Lorincz et al., 2012; Deruyck et al., 2012; Gadze et al., 2016; Ahn et al., 2019; Dahal, 2022; Chen et al., 2023. Linear regressions between power ratings and traffic have been generated but translating these into comparable equations for MEC base stations is impossible because the Erlang unit of telco traffic is ambiguous Iversen (2012); for example, 1 erl might equal 1 call lasting 60 min or 60 calls each lasting 1 min. Data for power ratings and traffic presented in Dahal (2022) offer, however, estimates of minimum (zero traffic) and full usage (maximum erl traffic) power ratings: 1.225 kW and 1.49 kW, respectively. From these data, a regression relationship was generated:

$$\mathrm{PR} = \mathrm{MPR}\left(1 + 0.1766 \times \left(\frac{\%\,\text{Full Usage}}{100}\right)\right) \tag{1}$$

With the MEC base station power estimate Dongxu (2020), this becomes:

$$\mathrm{PR} = 11577\left(1 + 0.1766 \times \left(\frac{\%\,\text{Server CPU Usage}}{100}\right)\right) \tag{2}$$

For 2/3 G and 4 G base stations, the MPR value is lower Dongxu (2020). It is our assumption that linear relationships exist for all the generations of base stations and their workload as quantified by the CPU workload of the linked server (Table 2).

For a base station-linked server, data for 6 data centre servers were used from published data Ji et al. (2021) and reordered into increasing peak server power (Table 2). Types D and F originally had identical CPU core numbers but this was altered to more clearly differentiate server types. Type F had idiosyncratically high power per task values when CPU usage was low because of the power rating values selected Ji et al. (2021).

The non-linear increase in power rating for a server with increasing CPU utilisation (workload) was taken from Fan et al. (2007). Server power was multiplied by a factor of 1.5 to estimate the total server plus ancillary (mostly cooling system) power consumption as deduced from analysis of real-world data sets Ji et al. (2021). Each CPU core was assumed to be capable of handling 250 jobs at any one time Smith (2022). To model increasing server power consumption with simultaneous task numbers, linear models between idle and peak power values were constructed and related to actual CPU usage and jobs/tasks being processed as CPU usage increased, the power consumed per task decreased (Figs. 2 and 3).

Power rating values for a MD were taken from Kumar and Lu (2010): $P_{\text{transmit}} = P_{\text{receive}} = 0.9\,\text{W}$, $P_{\text{compute}} = 1.3\,\text{W}$ and $P_{\text{idle}} = 0.3\,\text{W}$. The P_{transmit} estimate (0.9 W) is also a value calculated from the analysis of 545 million devices operating in commercial 5 G networks Joshi et al. (2020).

The MD CPU workload was initially minimised at 0% to bias the energy balance in favour of local computation. The MD and server CPU processing speeds were taken from data in Melendez and McGarry (2017): MD processor $= 3.6 \times 10^9$ instructions per second (ips),

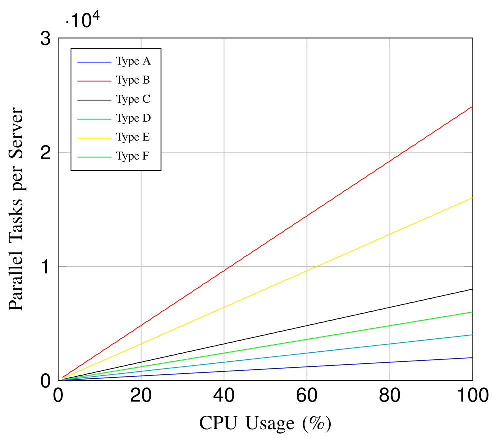

Fig. 2. Increased parallel task activities with increased CPU usage in six types of server computed from data in Ji et al. (2021).

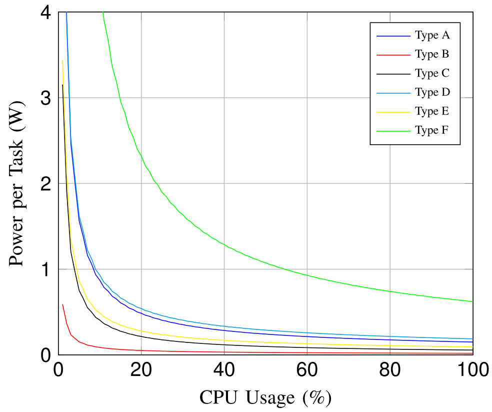

Fig. 3. Trends in power per unit task with increased CPU usage in six types of server computed from data in Ji et al. (2021).

MEC server processor = 1.4×10^{11} ips. Increasing CPU usage decreased these speeds as computed in Singh et al. (2020b). The baseline scenario used a 4 MB file offloaded from the MD to a unique edge server connected to a base station with no latency or queuing delays. Returned data was 10% of that transmitted. The lowest complexity app from Melendez and McGarry (2017) was taken as 0.0277 bpi (bits per instruction). This was used in the baseline scenario to compute energy usage in the MD or in the MD-edge computing dialogue. The app was one of 9 scientific software programs whose bpi values were computed from workload data in a computing centre Melendez and McGarry (2017). The authors of Melendez and McGarry (2017) used bpi values to relate file size (bits) to the size of the computational task (instructions); the smaller the bpi value, the more complex was the computational task.

The total energy consumption for offloading was made up from energy use by the MD and Edge Computing hardware components, respectively:

$$[T_T \times P_{TR} + T_C \times P_I + T_R \times P_{TR}] + [(E_{BS}/_n) \times (T_T + T_R] + P_S \times T_C \tag{3}$$

where T_T (s) is the data transmission over the WLAN, T_C (s) is the computation time on the edge node server, T_R (s) is the data return time from the edge node, P_{TR} (W) is the MD power rating while transmitting or receiving and P_I (W) is the MD power rating when idling (i.e., during T_C) P_S (W) is the power rating per task of a base station server at a known CPU workload, and E_{BS} (W) is the total base station power consumption per task at any individual CPU workload value (equation (2)).

The total energy consumption for local data processing on the MD was $T_{MD} \times P_C$, where T_{MD} was the time required for computing on the MD and P_C was the MD power rating while computing. T_{MD} was computed as: $(X/CC_T)/S_{MD}$, where (X) was the size of the computational task (bits), CC_T was the computational complexity (bpi) of the task and S_{MD} was the on-board processing speed (ips) of the MD.

5. Performance evaluation

This section discusses the experimental results from numerical simulations. We have considered different WLAN speeds for our experiments, such as 50 Mbps WLAN and 500 Mbps WLAN.

5.1. 50 Mbps WLAN

Using the input data as described in Section 4, total energy usage per offloaded task was highly variable depending on the exact base station-server combination, with the worst case scenario (type A server) energy being > 12 times that of local computation (Fig. 4). With other base-station-server combinations, however, offloading was relatively energy-efficient with three of the six server types when the MEC server CPU usage reached 50% (Fig. 4). No base station server type showed any total energy usage reduction when the server CPU workload was only 10% (Table 3).

The type B server showed energy savings from offloading at a 20% server CPU workload; type E showed energy savings at a 30% server CPU workload; type C showed energy savings at 50% server CPU workloads. In contrast, types A, D and F showed no energy savings at any of the investigated server workloads (Table 3). The maximum energy saving over local computation was 51% in the type B server when the server CPU workload was 50%.

At 50% server CPU workload, the majority (89%) of the total MEC energy use was in the relatively lengthy data reception phase (Fig. 5). The data re-transmission phase accounted for 8.9% of the base station-server edge node on average while the computation phase only accounted for, on average, 2.2% of the total edge node energy usage.

With a 4 G base station and its lower power rating Dongxu (2020), the maximum energy savings by offloading reached 60% with a type B server at 50% server CPU workload. The data transmission time from the MD to a base station is much longer than during processed data return back to the MD; this is because only 10% of the originally transmitted data was returned as processed results. This is the same for all MD-base station combinations (Fig. 6).

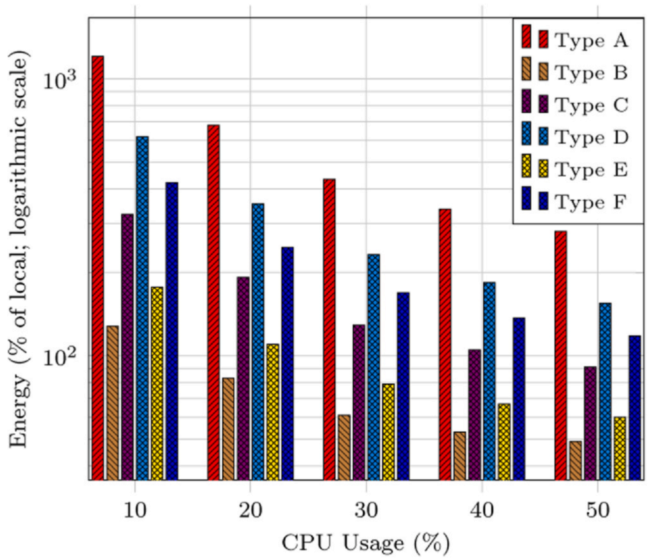

Fig. 4. Total energy use in offloading at a WLAN speed of 50 Mbps to a 5 G MEC base station connected to one of different server types Ji et al. (2021) all with CPU workloads of 50%.

Table 3
Total, edge node (base station and server) and local (MD) energy saving compared to MD use only with varying server CPU workloads, a 4MB data file and a WLAN speed of 50 Mbps. The "–" entry denotes more energy required for offloading than for local computation on the MD.

Base station configuration	Server CPU Workload 10%	Server CPU Workload 20%	Server CPU Workload 50%
Type A server maximum energy saving (%)	-	–	–
Type B server maximum energy saving (%)	–	17	47
Type C server maximum energy saving (%)	–	–	–
Type D server maximum energy saving (%)	–	–	–
Type E server maximum energy saving (%)	–	–	33
Type F server maximum energy saving (%)	–	–	–
MD energy (j)	3.7	3.7	3.7
Edge node minimum energy (J/task)	4.7	3.1	2.0
Edge node maximum energy (J/task)	44.6	25.1	12.5

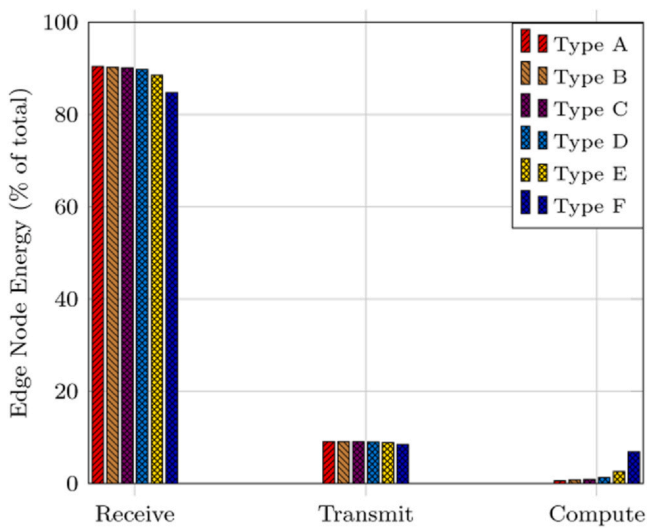

Fig. 5. Breakdown of energy use by the 5 G edge node base station and server into a MEC base station connected to one of six different server types Ji et al. (2021). WLAN: 50 Mbps; server CPU workload: 50%.

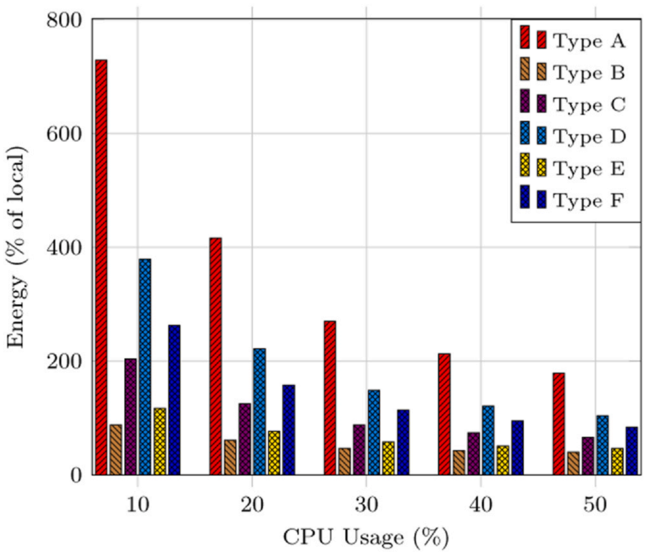

Fig. 6. Energy efficiency (relative to on-device computation) in offloading at a WLAN speed of 50 Mbps to a 4 G MEC base station connected to one of six different server types Ji et al. (2021). WLAN: 50Mbps; server CPU workload: 50%.

Type A and F were, however, unable to demonstrate any energy savings. With a 2/3 G base station, the maximum energy savings by offloading reached 63% with a type B server at 50% server CPU workload; only the edge node with the type A server configuration failed to demonstrate any energy savings in the 2/3 G-server edge node (Fig. 7). We interpret this as a consequence of the much lower power rating of the 2/3 G era base stations Dongxu (2020); in performance terms, however, 4 G and 5 G base stations achieve superior network data transmission times and shorter offloading times. The Type A had the smallest number of tasks processed in parallel and a relatively high peak power rating per task. The combination of these factors explains its sub-optimal performance characteristics.

The sufficiently high CPU workload of the base station-linked server was a critical parameter for achieving maximum energy savings by offloading and to achieve energy savings with a broader range of server types (Table 3). A server CPU workload as low as 10% eliminated any energy advantage of offloading with the 5 G base station. Only one server type exhibited any energy saving at a 10% server workload with the 4 G base station, and two server types achieved energy savings at the 10% server workload with the 2/3 G base station (Table 4).

The computational task complexity can be decreased by multiplying the app bpi value by a factor > 1.0; computational complexity is inversely proportional to the bpi value Melendez and McGarry (2017). When a factor of 2.2 was used (bpi $= 4.994 \times 10^{-3}$), the energy advantage of offloading also disappeared although the type B server option had total energy only 5%

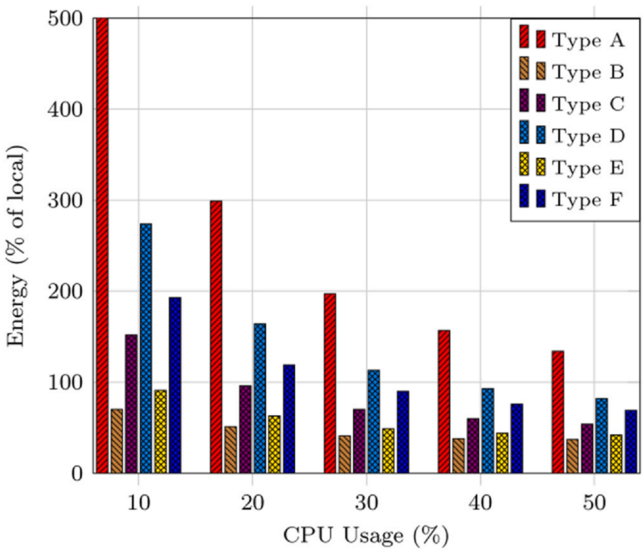

Fig. 7. Energy efficiency (relative to on-device computation) in offloading at a WLAN speed of 50 Mbps to a 3 G MEC base station connected to one of six different server types Ji et al. (2021), all with CPU workloads of 50%.

Table 4
Total, edge node (4 G and 2/3 G base station [BS] and server) and local (MD) energy saving compared to MD use only with varying server CPU workloads for a 4 G and 2/3 G base station, a 4MB data file and a WLAN speed of 50 Mbps. The "–" entry denotes more energy required for offloading than for local computation on the MD.

Base station configuration	4 G BS Server CPU Workload 10%	4 G BS Server CPU Workload 20%	2/3 G BS Server CPU Workload 10%
Type A server maximum energy saving (%)	–	–	–
Type B server maximum energy saving (%)	12	53	30
Type C server maximum energy saving (%)	–	–	–
Type D server maximum energy saving (%)	–	–	–
Type E server maximum energy saving (%)	–	42	9
Type F server maximum energy saving (%)	–	–	–
MD energy (j)	3.7	3.7	3.7
Edge node minimum energy (J/task)	3.25	1.75	2.6
Edge node maximum energy (J/task)	26.9	10.0	19.2

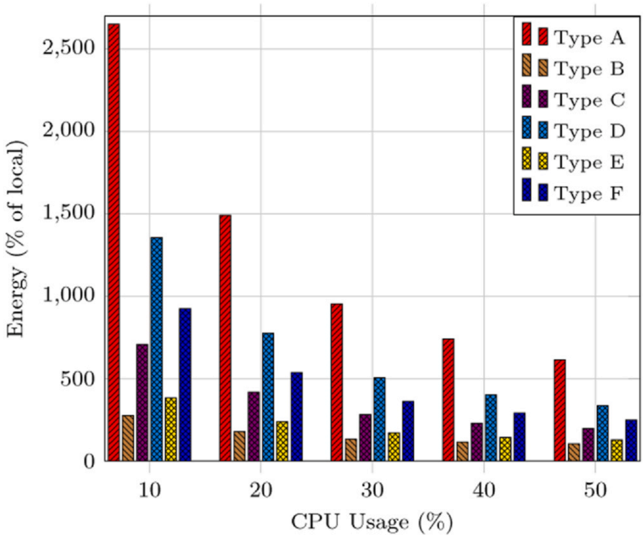

Fig. 8. Energy efficiency (relative to on-device computation) in offloading at a WLAN speed of 50 Mbps a task with a computational complexity of (4.99×10^{-3} bpi) to a 5 G MEC base station connected to one of six different server types Ji et al. (2021), all with CPU workloads of 50%.

greater than that of local computation (Fig. 8 and Table 5). Our interpretation is that relatively simple tasks require comparatively short computation local (MD) times and low on-board energy usage.

Conversely, increasing the computational complexity (by reducing the bpi value to 7.491×10^{-4}) resulted in total energy savings with five of the six server types at a server CPU workload of 50% (Fig. 9 and Table 5). Offloading the task had a maximum energy saving of approximately 80% with server types B and E when compared with local computation. Our interpretation is that complex computational tasks require longer computation local (MD) times and high on-board energy usage.

Table 5
Total, edge node (base station and server) and local (MD) energy saving compared to MD use only with server CPU workloads of 50%, a 4MB data file and a WLAN speed of 50 Mbps with varying task complexity and MD CPU workload (0% and 84%). The "–" entry denotes more energy required for offloading than for local computation on the MD.

Base station configuration	Task 1.794 bpi ($\times 10^{-4}$)	Task 7.491 bpi ($\times 10^{-4}$)	Task 4.994 bpi ($\times 10^{-3}$)	Task 4.994 bpi ($\times 10^{-3}$) MD CPU 84%
Type A server maximum energy saving (%)	75	–	–	2
Type B server maximum energy saving (%)	94	55	–	83
Type C server maximum energy saving (%)	91	–	–	68
Type D server maximum energy saving (%)	85	–	–	46
Type E server maximum energy saving (%)	93	39	–	79
Type F server maximum energy saving (%)	83	–	–	60
MD energy (j)	46.8	11.2	1.7	10.5
Edge node minimum energy (J/task)	2.6	5.0	1.8	1.8
Edge node maximum energy (J/task)	11.7	44.9	10.3	10.3

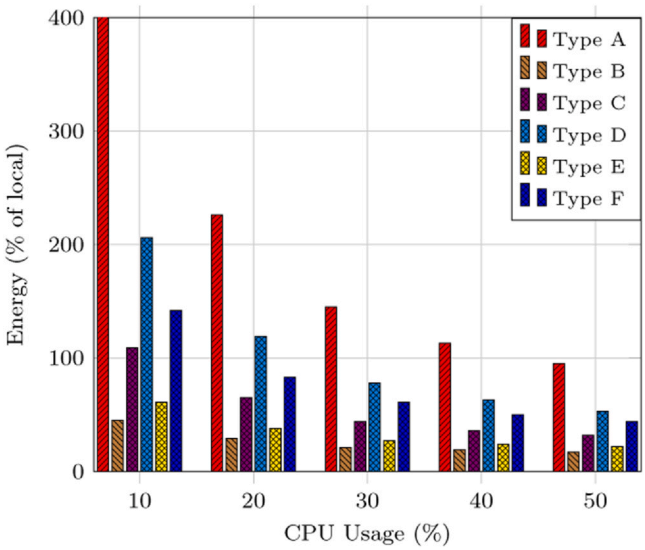

Fig. 9. Energy efficiency (relative to on-device computation) in offloading at a WLAN speed of 50 Mbps a task with higher computational complexity (7.49×10^{-4} bpi) to a 5 G MEC base station connected to one of six different server types Ji et al. (2021), all with CPU workloads of 50%.

Increasing the task complexity further (bpi = 1.793×10^{-4}), resulted in all server types being capable of showing total energy reductions even at a server workload of 10% (Fig. 10 and Table 5). Under these conditions, the maximum energy saving by offloading to the edge node was 94%, with server type B at a 50% server workload. All server types gave a minimum of 75% energy saving under these conditions.

A high CPU usage in the MD (84%) significantly increased local computation time and energy use; this resulted in offloading, giving greater energy savings over local computation with all base station-server combinations even with low-complexity apps (Fig. 11 and Table 5). The maximum savings of approximately 80% over local computation was achieved with server types B. and E.

Increasing the MD task size to 10 MB or decreasing the task to 1 MB yielded identical results to a 4 MB task. Similarly, task partitioning between multiple servers in the base station could not increase total energy savings; although shorter computation times were possible in the edge network, the summation of multiple server energies eliminated any energy savings advantages.

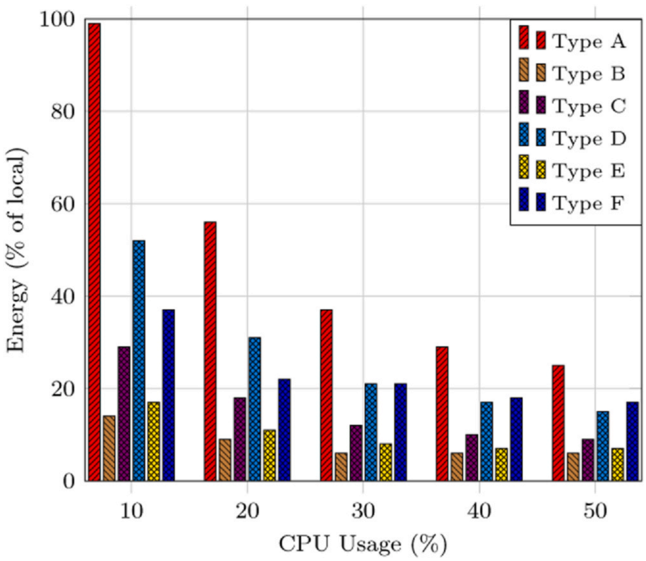

Fig. 10. Energy efficiency (relative to on-device computation) in offloading at a WLAN speed of 50 Mbps a task with the highest computational complexity (1.79×10^{-4} bpi) to a 5 G MEC base station connected to one of six different server types Ji et al. (2021), all with CPU workloads of 50%.

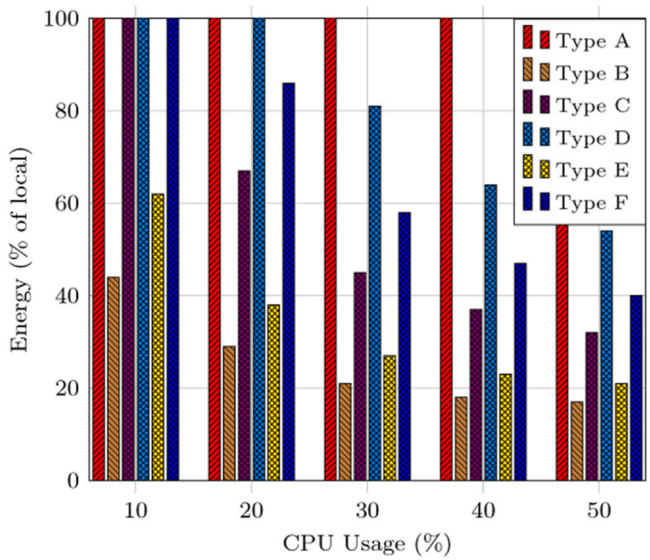

Fig. 11. Energy efficiency (relative to on-device computation) in offloading at a WLAN speed of 50 Mbps (for the WLAN speed) a task with higher computational complexity (1.793×10^{-4} bpi) to a 5 G MEC base station connected to one of six different server types Ji et al. (2021), all with CPU workloads of 50%.

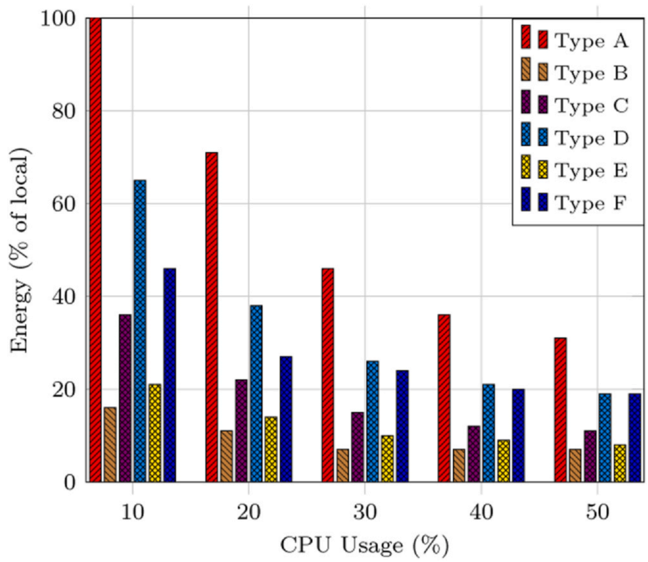

Fig. 12. Total energy use in offloading at a WLAN speed of 500 Mbps to a 5 G MEC base station connected to one of six different server types Ji et al. (2021).

5.2. 500 Mbps WLAN

With a ten-fold higher WLAN speed, all types of base station-server combinations could deliver energy savings by offloading a 4MB data file at moderate (20% CPU workloads (Fig. 12). The maximum energy savings were 93% (type B, 20% CPU workload) and 92% (type E, 50% CPU workload). With increasing WLAN speed, the energy savings over local computation increased rapidly between 100 and 300 Mbps (Fig. 13). The energy used by the edge node (base station plus server) decreased rapidly in this range of WLAN speeds as progressively less edge node energy was expended during the reception phase of the MD data (Fig. 14). As with a lower (50 Mbps) WLAN speed, energy savings as a percentage of local (MD) computing for data files between 1 MB and 100 MB were identical even though absolute energy use (local and edge node) increased with increasing data file size.

The use of the higher WLAN speed to the base node in the MEC network allowed less complex computational tasks to be offloaded with energy savings; the savings were greater at a 50% CPU workload than at a 10% CPU workload (Table 6). Across all 6 server types

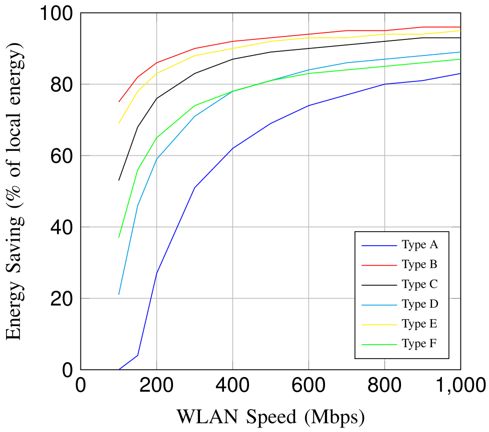

Fig. 13. Energy efficiency (relative to on-device computation) in offloading a 4MB task at increasing WLAN speeds to a 5 G MEC base station connected to one of six different server types Ji et al. (2021) with 50% CPU workloads.

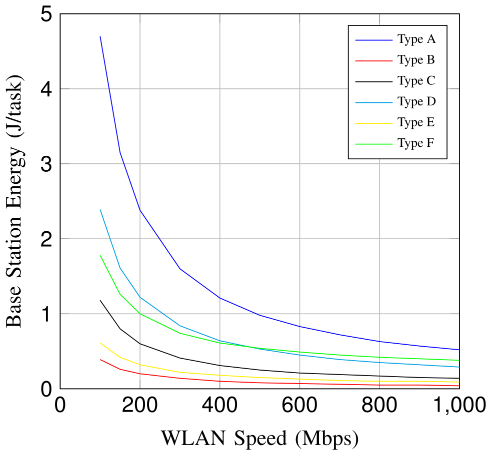

Fig. 14. Edge node (base station + server) energy use in offloading a 4 MB task at increasing WLAN speeds to servers with 50% CPU workloads.

included in the analysis, the optimal edge node server CPU workload for maximal total energy savings was approximately 50% (Fig. 15). Above a 90% CPU workload, a severe deterioration in total energy savings began to occur due to much longer server computation times for offloaded tasks.

A high (84%) MD CPU workload resulted in very high energy savings with all base station-server combinations at server CPU workloads of 50% (Table 6).

Table 6
Total, edge node (base station and server) and local (MD) energy saving compared to MD use only with server CPU workloads of 50%, a 4MB data file and a WLAN speed of 500 Mbps with varying task complexity and MD CPU workload (0% and 84%). The "–" entry denotes more energy required for offloading than for local computation on the MD.

Base station configuration	Task 4.994 bpi ($\times$ 10^{-3})	Task 1.135 bpi ($\times$ 10^{-2})	Task 1.7 bpi ($\times$ 10^{-2})	Task 2.27 bpi ($\times$ 10^{-3}) MD CPU 84%
Type A server maximum energy saving (%)	36	–	–	95
Type B server maximum energy saving (%)	88	75	63	99
Type C server maximum energy saving (%)	78	53	31	98
Type D server maximum energy saving (%)	63	21	–	97
Type E server maximum energy saving (%)	85	69	54	99
Type F server maximum energy saving (%)	68	37	10	97
MD energy (j)	1.7	0.7	0.5	23.1
Edge node minimum energy (J/task)	0.1	0.2	0.2	0.2
Edge node maximum energy (J/task)	1.1	1.1	1.0	1.1

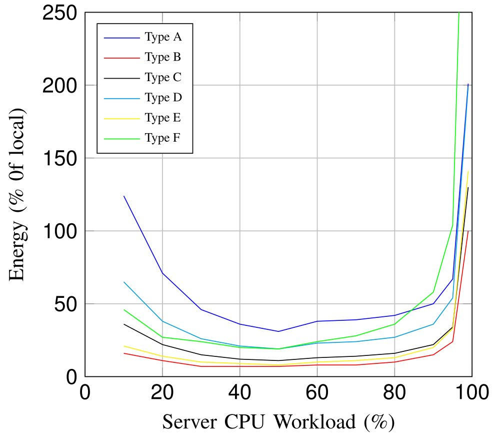

Fig. 15. Effect of edge node server CPU workload on total energy savings by offloading with 6 different server types.

6. Conclusions and future work

Our work has demonstrated the large potential energy savings (compared to the energy usage of an individual MD) by offloading tasks to an edge node server capable of processing large numbers of tasks with a significantly leveraged processor speed. Achieving this "economy of scale"İ is a logical outcome of the highly desirable features of edge resources, i.e., superior resources based on high-capacity servers to minimise energy per completed task, which our analysis has demonstrated to be the critical parameter for purposes of comparison. This will make mobile edge computing services more sustainable and environmentally friendly as an initial step towards net zero emissions to achieve carbon neutrality.

We acknowledge that to reach these quantitative conclusions, we have deliberately simplified some features of the analysis. In particular, we assume that an SLA between the individual user for a premium service eliminates any significant latency in the MD-Edge computing network linkage. In practice, such a complete elimination of latency and queuing delays may prove impossible; we intend to widen our analysis to explore limitations on offloading success in reducing energy demand in future work.

It is presently impossible to predict the full benefits to MD users and other enterprise stakeholders from reducing total energy demand in offloading. Extending MD battery lifetime is a major target, and the results from a 500Mbps 5 G base station link are highly promising in this regard. The only extra cost to 5 G Edge Computing services will be the development and installation of edge node software; this is analogous to "orchestration"İ programmes designed to maximise the efficiency of edge-cloud synergy (Singh and Kiss, 2022; Ullah et al., 2021. The implementation of such software with permanent access to the parameters of the MD-edge node relationship (processing speeds, app complexities, CPU workloads, etc.) will extend the service options available to the providers of Edge Computing networks and their many possible applications, the most pertinent of which to our study is computational offloading. Viewed more broadly, this approach could be extended to Augmented Reality applications hosted by Edge Computing networks to increase commercial viability and service uptake by users.

We recommend that to achieve the maximum possible energy savings for MD users from offloading service providers, invest in the most computationally adept servers, i.e., those with high numbers of cores and maximal numbers of jobs capable of simultaneous processing while minimising their idling and peak server power ratings.

From the standpoint of the individual MD user, high energy efficiency per task requires edge node CPU workloads to be sufficiently high that the energy per completed task is minimised. This requirement is consistent with not offloading to the MEC network when user demand is low, although enterprise traffic for Big Data applications could even out daily workloads even at off-peak times. Conversely, times of the highest demand for the MEC resources also reduce or even eliminate any potential overall energy savings. The actual energy savings by offloading also depend critically on the precise configuration of the edge node hardware; this information may not be available to any user of MDs connecting to the MEC network. It is likely, however, that the evolution of future MEC resources will be paralleled by a greater user appreciation of resource availability and its variability; the use of Artificial Intelligence (AI) software techniques in decision-making for offloading events may be a practical path to matching user demand, offloaded task completion time and overall energy efficiency Carvalho et al. (2020). For example, AI embedded in the edge node could communicate to the MD users both the actual energy (and computational time) savings in real-time and forecast short-term trends in these offloading parameters (Walia et al., 2023; Gill et al., 2024a.

Digital communication devices - and, in particular, smartphones - pose severe challenges to Circular Economy principles because of consumer practice as well as hardware and technology factors Cordella et al. (2021). Our results demonstrate that MEC networks offer means of reducing total energy usage in computation offloading. The impact on global carbon emissions is less clear because both the networks and MD users may not be able to utilise fully carbon-neutral electricity for some decades and progress towards this goal will be variable in different regions of the global economy. This quantitative analysis of energy-efficient edge computing would be useful for achieving carbon neutrality and contributing towards the initial net zero emission targets. In the future, we will explore if fixed-location computers could reduce energy usage as well as benefit from shorter computational task times by offloading to edge node resources Singh and Kiss (2022) and from edge-cloud synergies in data processing using AI and measure the impact of energy consumption on carbon emissions Gill et al. (2024a).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahn, J., Lee, J., Park, S., Park, H.-S., 2019. Power efficient clustering scheme for 5G mobile edge computing environment. Mob. Net. Appl. 24, 643–652.

Andrae, A.S., Edler, T., 2015. On global electricity usage of communication technology: trends to 2030. Challenges 6 (1), 117–157.

O. Arnold, F. Richter, G. Fettweis, and O. Blume, Power consumption modeling of different base station types in heterogeneous cellular networks, In: Proceedings of the Future Network & Mobile Summit, Florence, Italy, 2010, 1–8.

Cao, Z., Zhou, X., Hu, H., Wang, Z., Wen, Y., 2022. Toward a systematic survey for carbon neutral data centers. IEEE Commun. Surv. Tutor. 24 (2), 895–936.

Carvalho, G., Cabral, B., Pereira, V., Bernardino, J., 2020. Computation offloading in edge computing environments using artificial intelligence techniques. Eng. Appl. Artif. Intell. 95, 103840.

Chen, Y., Xun, L., Zhang, S., 2023. The energy efficiency of heterogeneous cellular networks based on the poisson hole process. Future Internet 15 (2), 56.

K. Cheng, Y. Teng, W. Sun, A. Liu, and X. Wang, Energy-efficient joint offloading and wireless resource allocation strategy in multi-mec server systems, In: Proceeding of the 2018 International Conference on Communication, IEEE, 2018, Kansas City, 1–6.

S. Clinch, J. Harkes, A. Friday, N. Davies, and M. Satyanarayanan, How close is close enough? understanding the role of cloudlets in supporting display appropriation by mobile users, In: IEEE International Conference on Pervasive Computing and Communications (PerCom), Lugano, Switzerland, 2012, 122–127.

Cordella, M., Alfieri, F., Clemm, C., Berwald, A., 2021. Durability of smartphones: a technical analysis of reliability and repairability aspects. J. Clean. Prod. 286, 125388.

Dahal, M.S., 2022. Energy saving in 5G mobile communication through traffic driven cell zooming strategy. Energy Nexus 5, 100040.

Deruyck, M., Vereecken, W., Joseph, W., Lannoo, B., Pickavet, M., Martens, L., 2012. Reducing the power consumption in wireless access networks: overview and recommendations. Prog. Electro Res. 132, 255–274.

C. Dongxu.5G power: creating a green grid that slashes costs, emissions and energy use, 2020.[Online]. Available: https://www.huawei.com/uk/technology-insights/publications/huawei-tech/89/5g-power-green-grid-slashes-costs-emissions-energy-use.

Du, J., et al., 2023. Computation energy efficiency maximization for intelligent reflective surface-aided wireless powered mobile edge computing. IEEE Trans. Sustain. Comput.

Fan, X., Weber, W.-D., Barroso, L.A., 2007. Power provisioning for a warehouse-sized computer. Proc. 34th Annu. Int. Symp. . Comput. Archit., San. Diego, CA 35 (2), 13–23.

R. Fantacci and B. Picano, Performance analysis of an edge computing system for real time computations and mobile users, In: 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China, 2020, 1540–1545.

Gadze, J.D., Aboagye, S.B., Agyekum, K.A., 2016. Real time traffic base station power consumption model for telcos in ghana. Int. J. Comp. Sci. Telecom 7 (5), 6–13.

Gill, S.S., et al., 2024a. Modern computing: vision and challenges. Telemat. Inform. Rep. 1–46.

Gill, S.S., et al., 2024b. Transformative effects of ChatGPT on modern education: emerging era of AI chatbots. Internet Things Cyber-Phys. Syst. 4, 19–23.

Guo, S., Liu, J., Yang, Y., Xiao, B., Li, Z., 2019. Energy-efficient dynamic computation offloading and cooperative task scheduling in mobile cloud computing. IEEE Trans. Mob. Comput. 18, 319–333.

W. Hu, Y. Gao, K. Ha, J. Wang, B. Amos, Z. Chen, P. Pillai, and M. Satyanarayanan, Quantifying the impact of edge computing on mobile applications, In: Proceedings of the 7th ACM SIGOPS Asia-Pacific Workshop on Systems, Hong Kong, 2016, 1–8.

Iftikhar, S., et al., 2022. Ai-based fog and edge computing: a systematic review, taxonomy and future directions. Internet Things 100674.

V.B. Iversen, The internet erlang formula, In: Internet of Things, Smart Spaces, and Next Generation Networking. ruSMART NEW2AN 2012 2012.Lecture Notes in Computer Science, vol. 7469. Springer, Berlin, Heidelberg.

Ji, K., Chi, C., Zhang, F., Anta, A.F., Song, P., Marahatta, A., Wang, Y., Liu, Z., 2021. Energy-aware scheduling based on marginal cost and task classification in heterogeneous data centers. Energies 14 (9), 2382.

Jiang, C., Cheng, X., Gao, H., Zhou, X., Wan, J., 2019. Toward computation offloading in edge computing: a survey. IEEE Access 7, 131543–131558.

Joshi, P., Ghasemifard, F., Colombi, D., Törnevik, C., 2020. Actual output power levels of user equipment in 5g commercial networks and implications on realistic rf emf exposure assessment. IEEE Access.

Kumar, K., Lu, Y.-H., 2010. Cloud computing for mobile users: can offloading computation save energy? Computer 43 (4), 51–56.

Kurniawan, T.A., Othman, M.H.D., Liang, X., Goh, H.H., Gikas, P., Kusworo, T.D., Anouzla, A., Chew, K.W., 2023. Decarbonization in waste recycling industry using digitalization to promote net-zero emissions and its implications on sustainability. J. Environ. Manag. 338, 117765.

K. Lin, B. Liu, Y. Chen, H. Dai, and Z. Liang, Intelligent edge computing gateway implementation method based on microservice architecture, In: 2022 International Conference on Informatics, Networking and Computing (ICINC). IEEE, 2022, 203–208.

Lin, L., Liao, X., Jin, H., Li, P., 2019. Computation offloading toward edge computing. Proc. IEEE 107 (8), 1584–1607.

Liu, J., Shou, G., Liu, Y., Hu, Y., Guo, Z., 2018. Performance evaluation of integrated multi-access edge computing and fiber-wireless access networks. IEEE Access 6, 30269–30279.

L. Liu, Z. Chang, X. Guo, and T. Ristaniemi, Multi-objective optimization for computation offloading in mobile-edge computing, In: IEEE Symposium on Computers and Communications (ISCC), IEEE, 2017, 832–837.

Lorincz, J., Garma, T., Petrovic, G., 2012. Measurements and modelling of base station power consumption under real traffic loads. Sensors 12 (4), 4181–4310.

M.P. Mahenge, C. Li, and C.A. Sanga, Collaborative mobile edge and cloud computing: Tasks unloading for improving users' quality of experience in resource-intensive mobile applications, In: 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, 322–326.

Mahenge, M.P.J., Li, C., Sanga, C.A., 2022. Energy-efficient task offloading strategy in mobile edge computing for resource-intensive mobile applications. Digit. Comm. Netw. 8, 1048–1058.

F. Malandrino, S. Kirkpatrick, and C.-F. Chiasserini, How close to the edge? Delay/utilization trends in MEC, In: Proceedings of the ACM Workshop on Cloud-Assisted Networking, 2016, 37–42.

Y. Mao, J. Zhang, S. Song, and K.B. Letaief, Power-delay tradeoff in multi-user mobile-edge computing systems, In: 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA, 2016, 1–6.

S. Melendez and M.P. McGarry, Computation offloading decisions for reducing completion time, In: Proceedings of the 14th IEEE Annual Consumer Communications Networking Conference (CCNC), 2017, 160–164.

Nandhakumar, A.R., et al., 2023. EdgeAISim: a toolkit for simulation and modelling of AI models in edge computing environments. Meas.: Sens.

Ramaraj, S., 2016. Tablet computers in education. J. Contemp. Ed. Res. Innov. 6 (6), 258–262.

Sabella, D., Vaillant, A., Kuure, P., Rauschenbach, U., Giust, F., 2016. Mobile-edge computing architecture: the role of MEC in the internet of things. IEEE Consum. Electron. Mag. 5 (4), 84–91.

Sabella, D., Reznik, A., Frazao, R., 2020. Multi-access edge computing in action. CRC Press.

Safavat, S., Sapavath, N.N., Rawat, D.B., 2020. Recent advances in mobile edge computing and content caching. Digit. Comm. Net. 6 (2), 189–194.
Senthilkumar, A., Prabhu, L., Sathish, T., Saravanan, R., Jeyaseelan, G.A.C., Ağbulut, Ü., Mahmoud, Z., Shaik, S., Saleel, C.A., 2023. Enhancement of r600a vapour compression refrigeration system with mwcnt/tio2 hybrid nano lubricants for net zero emissions building. Sustain. Energy Technol. Assess. 56, 103055.
Seto, K.C., Churkina, G., Hsu, A., Keller, M., Newman, P.W., Qin, B., Ramaswami, A., 2021. From low-to net-zero carbon cities: The next global agenda. Annu. Rev. Environ. Resour. 46, 377–415.
R. Singh et al., Edge ai: a survey, Internet of Things and Cyber-Physical Systems, 2023.
R. Singh and T. Kiss, Edge-cloud synergy: Unleashing the potential of parallel processing for big data analytics, In: in IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, Canada, 2022, 0001–0006.
R. Singh, S. Armour, A. Khan, M. Sooriyabandara, and G. Oikonomou, Heuristic approaches for computational offloading in multi-access edge computing networks, In: IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC): Track 4: Applications and Business, 2020a, 1–6.
R. Singh, S. Armour, A. Khan, M. Sooriyabandara, and G. Oikonomou, Identification of the key parameters for computational offloading in multi-access edge computing, In: Proceedings of IEEE Cloud Summit, 2020b, 131–136.
Singh, R., Armour, S., Khan, A., Sooriyabandara, M., Oikonomou, G., 2019. The advantage of computation offloading in multi-access edge computing. 2019 Fourth International Conference on Fog and Mobile Edge Computing (FMEC). pp. 289–294.
R. Singh, S. Armour, A. Khan, M. Sooriyabandara, and G. Oikonomou, Towards multi-criteria heuristic optimization for computational offloading in multi-access edge computing, In: 2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR), Paris, France, 2021, 1–6.
M. Smith.How many concurrent requests can a web server handle? 2022.(Accessed 6 April 2023).[Online]. Available: https://developerpitstop.com/how-many-concurrent-requests-can-a-web-server-handle/#:~:text=With%20a%20single%20CPU%20core,your%20site%20grows%20in%20popularity.
C. Sonmez, A. Ozgovde, and C. Ersoy, Edgecloudsim: An environment for performance evaluation of edge computing systems, 2017 Second International Conference on Fog and Mobile Edge Computing (FMEC), Valencia, Spain, 2017, 39–44.
Ullah, A., Dagdeviren, H., Ariyattu, R.C., DesLauriers, J., Kiss, T., Bowden, J., 2021. Towards an application-level orchestrator for the cloud-to-edge computing continuum. J. Grid Comput. 19, 1–28.
B. Varghese, N. Wang, S. Barbhuiya, P. Kilpatrick, and D.S. Nikolopoulos, Challenges and opportunities in edge computing, In: 2016 IEEE international conference on smart cloud (SmartCloud) New York, NY, USA, 2016, 20–26.
Walia, G.K., et al., 2023. Ai-empowered fog/edge resource management for iot applications: a comprehensive review, research challenges and future perspectives. IEEE Commun. Surv. Tutor. 26 (1), 1–56.
Y. Wang, M. Tang, S. Zhou, G. Tan, Z. Zhang, and J. Zhan, Performance analysis of heterogeneous mobile edge computing networks with multi-core server, In: 2020 IEEE 20th International Conference on Communication Technology (ICCT). IEEE, 2020, 1540–1545.
Y. Wen, W. Zhang, and H. Luo, Energy-optimal mobile application execution: Taming resource-poor mobile devices with cloud clones, In: Proceeding of the 2012 IEEE INFOCOM, IEEE, Orlando, 2012, 2716–2720.
S. Yu and R. Langar, Collaborative computation offloading for multi-access edge computing, In: 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Arlington, VA, USA, 2019, 689–694.
J. Zhang and X. Zhao, An overview of user-oriented computation offloading in mobile edge computing, In: 2020 IEEE World Congress on Services (SERVICES), Beijing, China, 2020, 75–76.
Z. Zhuang, K.-H. Kim, and J.P. Singh, Improving energy efficiency of location sensing on smartphones, In: ACM SIGMOBILE International Conference on Mobile Systems, Applications, and Services, San Francisco, USA, 2010, 315–330.