Classification of optimization problems in fog computing[☆]

Julian Bellendorf[a,*], Zoltán Ádám Mann[a]

[a] *paluno – the Ruhr Institute for Software Technology, University of Duisburg-Essen, Essen, Germany*

Abstract

Fog computing combines cloud services with geographically distributed resources near the network edge to offer computational offloading possibilities to end devices, featuring low latency. Optimization of various metrics (latency, bandwidth, energy consumption etc.) plays a vital role in fog computing. We present the results of a literature review on optimization in fog computing, covering 280 papers. In particular, we propose a taxonomy of different optimization problems in fog computing, a categorization of the metrics used in constraints and objective functions, and a mapping study of the relevant literature.

Keywords: fog computing, edge computing, optimization

1. Introduction

Fog computing is a new paradigm for providing computational resources in a geographically distributed way [1]. Similarly to cloud computing, also fog computing offers the access to computational resources as a service, using similar service models as cloud computing. Also similar technologies (virtualization, containers etc.) are used to ensure efficient management and good utilization of the resources. In contrast to cloud computing, which is based on a few high-capacity data centers, fog computing uses a large set of widely distributed and heterogeneous resources with moderate capacity, called fog nodes [2]. The main advantage of fog computing over cloud computing is the proximity to end devices like sensors and actuators, smartphones, smart cameras, devices from the Internet of Things (IoT) etc. Such end devices typically have very limited capacities. Therefore, applications involving the end devices often require also more powerful devices. One possibility is to use the cloud as backend for such applications. However, this approach leads to high latency because of the long network path between end devices and cloud data centers; moreover, the bandwidth of the network connection of cloud data centers may become a major bottleneck. Fog computing solves these problems. Using fog nodes incurs significantly lower latency and eliminates the bottleneck on the cloud level [3].

Since its inception in 2012, fog computing has attracted an increasing amount of research. Optimization is very important already in cloud computing [4], and it plays an even more important role in fog computing research, since the fundamental goals of fog computing are related to optimization of metrics like latency, energy consumption, or resource utilization. Many researchers have formulated different problems in the design, deployment, and operation of fog computing systems as optimization problems. Different algorithmic techniques have been applied to solve those optimization problems [5]. The state of the art in optimization in fog computing is characterized by a proliferation of a large number of similar approaches, which are nevertheless different in their details. As a result, it is difficult to tell what the most important or mostly studied optimization problems are, to identify and compare different algorithms solving the same problem, or to adapt algorithms from one problem variant to another.

There are already some surveys on fog computing in general [6, 7, 8, 9, 10], on specific aspects of fog computing like security [11, 12, 13], or specific problems within fog computing like offloading [14, 15]. However, none of these surveys focuses on optimization explicitly and covers optimization in fog computing comprehensively.

[☆] This paper has been published in *Future Generation Computer Systems*, volume 107, pages 158-176, 2020.
[*] Corresponding author
Email address: julian.bellendorf@paluno.uni-due.de (Julian Bellendorf)

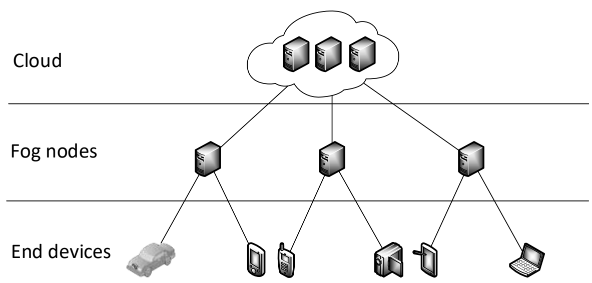


Figure 1: Architecture of fog computing

Therefore, the aim of this paper is to structure the existing research efforts on optimization in fog computing. For this purpose, we surveyed a total of 280 papers that propose optimization approaches in fog computing. We distilled the specific optimization problems solved by the papers, and organized these problems into a taxonomy with 4 categories and 8 subcategories. Moreover, we analyzed in detail the metrics used in the different papers for defining the objectives and the constraints of optimization problems. At first sight, just a handful of metrics are used in all papers. However, a more detailed investigation reveals that authors use the same term with many different meanings. For example, we identified 6 different metrics that are all regularly called "latency". We created a glossary of altogether 43 different metrics that play a role in optimizing fog computing systems. This allowed us a fine-grained categorization of the existing literature according to the defined problem variants and the used metrics. Our proposed categorization can be used to define canonical versions of the key optimization problems in fog computing, which paves the way to standardized benchmarks and meaningful comparison of different algorithms, ultimately contributing to the maturation of the fog computing field.

The rest of the paper is organized as follows. Section 2 introduces the necessary background in fog computing and in optimization. Section 3 describes the methodology followed in this work. Section 4 presents the taxonomy that we have elaborated. Section 5 provides a uniform definition of metrics. Section 6 shows in detail how existing work on optimization in fog computing could be categorized according to the presented taxonomy and metrics. Section 7 analyzes the results of the categorization, and Section 8 concludes the paper.

2. Preliminaries

2.1. Fog computing

We first review the fundamental properties of fog computing. For more details, we refer to the available general surveys on fog computing [6, 7, 8, 9, 10].

Fog computing supplies a distributed, heterogeneous infrastructure. As shown in Fig. 1, this infrastructure consists of three layers:

- **End devices** may include IoT devices, smartphones, sensors, actuators, smart cameras, connected cars, laptops etc. End devices can be numerous, heterogeneous, and mobile. Typically, the capacity of end devices, in terms of CPU and memory, is very limited. Often, end devices are battery-powered, leading to a limited battery lifetime and to the need to minimize power consumption. Often, end devices generate data that need to be processed and/or stored somewhere else (e.g., because the end device does not have sufficient capacity for processing or storing the data, or because data from multiple end devices have to be aggregated). In other cases, end devices need to receive data from somewhere else. Therefore, end devices are usually connected to some network (although continuous connectivity may not always be assumable).

- **Fog nodes** are computational resources available near the edge of the network, providing compute and storage services. Existing network equipment – e.g., routers, gateways, switches, base stations – with enough spare resources can act as fog nodes. In addition, a fog service provider may deploy resources (individual servers or small data centers) specifically for the purpose of acting as fog nodes. The capacity of fog nodes is typically higher than the capacity of end devices, but still limited. Fog nodes are geographically widely distributed, so that end devices can always connect to a near fog node (within a low number of network hops).

- **Cloud services** offer virtually unlimited capacity. The cloud services are provided by a small number of central large data centers, which may be far away from the end devices. Cloud services may be optionally used in fog computing.

To overcome their capacity limitations, end devices may use the services offered by fog nodes. For example, if an end device generates data that must be processed, but the capacity of the end device makes it impossible or impractical to process the data in the end device, the end device can connect to a nearby fog node, send the data to the fog node, and offload the processing of the data to the fog node. The low latency in network transfer between the end device and the fog node, coupled with the relatively high processing capacity of the fog node, make the whole process fast, allowing real-time interactions. This is important for many applications, for example in the road traffic management, augmented reality, and gaming domains. In such cases, offloading to the cloud would take too long. Nevertheless, cloud services can still be used, e.g., for storing aggregated data for later analysis [16].

Because of a lack of uniform terminology, some authors use other terms, like "edge computing", "mist computing", or "dew computing" to refer to this or similar concepts. In this paper, we always use the term "fog computing" to denote the described interplay between end devices, fog nodes, and potentially cloud services.

2.2. *Optimization problems*

An *optimization problem* typically consists of the following [17]:

- A set of variables that encode decisions to be made
- The set of possible values for each variable
- A set of constraints that the variables must satisfy
- An objective function

Both the constraints and the objective function may contain *metrics*, i.e., some numeric characteristics of the solution (e.g., latency or energy consumption). If $f(x_1, \ldots, x_n)$ is a metric depending on the value of variables $x_1, \ldots, x_n$, then $f(x_1, \ldots, x_n)$ may occur in the objective function (e.g., the objective function may be the weighted sum of several metrics), or in a constraint (e.g., in the form $f(x_1, \ldots, x_n) \leq C$ where C is a given constant).

A *solution* assigns to each variable one of its possible values, such that all constraints are satisfied. The aim of optimization is to find a solution that minimizes / maximizes the objective function.

3. Methodology

We followed a systematic methodology, which is summarized in Fig. 2. The first phase aimed at finding the set of relevant publications. This was challenging because (i) publications on fog computing are scattered over a large variety of journals and conferences and (ii) there is no uniform terminology on fog computing that would allow a simple keyword search. For these reasons, we applied a combination of three different search techniques:

- **Manual search**. We identified five conferences that deal specifically with fog computing. We manually checked each paper that was published in these conferences for its relevance to our work (see below for inclusion and exclusion criteria). The conferences are:
  - IEEE International Conference on Fog and Mobile Edge Computing (FMEC, since 2016)
  - IEEE International Conference on Fog and Edge Computing (ICFEC, since 2017)
  - IEEE International Conference on Edge Computing (EDGE, since 2017)
  - ACM/IEEE Symposium on Edge Computing (SEC, since 2016)

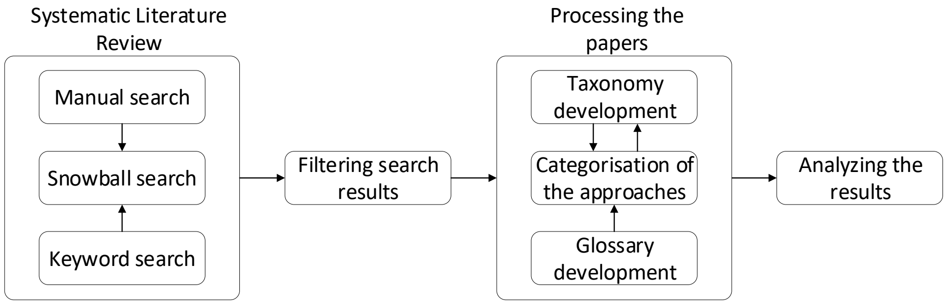


Figure 2: Overview of the used survey methodology

  - IEEE Fog World Congress (FWC, since 2017)

- **Keyword search**. On the basis of the identified relevant papers, we developed a search string. We started with the simple search string from [5] and extended it iteratively so that it matched at least 85% of the manually found papers (the original search string matched only about 20%). The resulting search string, which we applied to the Scopus[1] database, is:
  ```
  ((TITLE-ABS-KEY("edge computing" OR "fog computing" OR cloudlet)) OR (TITLE-ABS-KEY(offload*) AND
  TITLE-ABS-KEY(cloud))) AND (TITLE-ABS-KEY(optim* OR minimize OR maximize OR "objective function"))
  ```

- **Snowball search**. For each found paper, we checked the papers cited by or citing this paper to find further relevant papers.

The search was performed with a cutoff date of November 1st, 2018. The result of the combined search strategy was a set of potentially relevant papers that we further filtered according to the following criteria:

- We only included papers that deal explicitly with optimization problems in fog computing. This means that we excluded papers that are not about fog computing (e.g., papers in which tasks are offloaded from end devices to cloud services and not to fog nodes) as well as papers in which no explicit optimization problem is formulated (e.g., papers about technology and architecture issues in fog computing).

- We also excluded non-English papers as well as short papers (less than 4 pages in double-column format) that do not contain sufficient information to assess them.

During manual search, 285 papers were considered, 9 of which were selected as relevant for this work. Starting from these 9 papers, 40 further papers were found through snowball search. These 49 papers were used to define the search string. On our cutoff date the keyword search yielded about 1,700 papers, of which more than 1,420 were irrelevant. (In particular, many papers had to be discarded that contain these keywords but do not describe an optimization problem.) Overall, we identified 280 relevant publications.

We then analyzed these papers and extracted key information about the addressed optimization problems. On the basis of the extracted information, we developed a taxonomy of the main optimization problems identified. Parallel to that, we categorized each paper according to the taxonomy. This was an iterative process: we started with open coding, from which the taxonomy was gradually developed. As more and more papers were categorized, also changes to the taxonomy became necessary (also resulting in a re-categorization of already processed papers). In addition to the taxonomy of optimization problem variants, we also developed a glossary of the different metrics used in the papers, and we also used this glossary for tagging the papers.

In the final step, we analyzed the elaborated categorization of the papers to derive insights on the focal points of the existing research.

[1]https://www.scopus.com

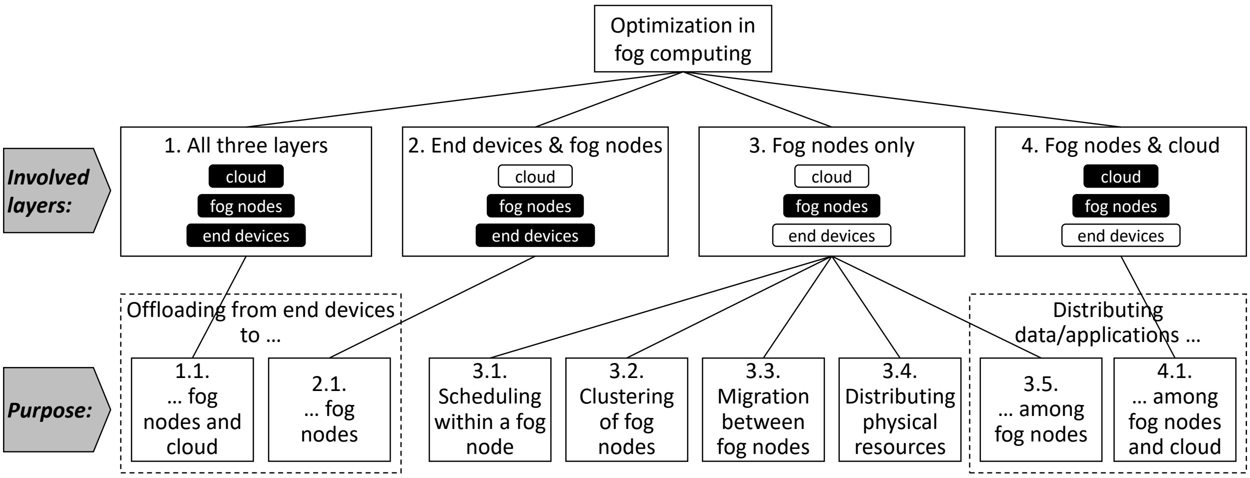


Figure 3: Taxonomy of optimization problems in fog computing

4. Taxonomy

This section explains the taxonomy of optimization problems that we developed for structuring the available literature on optimization in fog computing. The taxonomy, shown in Fig. 3, structures the literature into four categories and altogether eight subcategories.

The four categories are differentiated according to the involved architectural layers. As can be recalled from Fig. 1, the fog computing architecture includes three layers: end devices, fog nodes, cloud. Optimization problems in fog computing can relate to the following combinations of the three layers[2]:

1. All three layers. This combination occurs, for example, when computing tasks are offloaded from end devices and distributed over the available fog nodes and cloud services [18].

2. End devices and fog nodes. An example is the offloading of computing tasks from the end devices to the fog nodes [19].

3. Fog nodes only. Examples include decisions about migrating data between fog nodes [20].

4. Fog nodes and cloud. This combination occurs, for example, when data are distributed from the central cloud services over the fog nodes to make the data widely available [21].

Although other combinations of the architectural layers are also possible (e.g., cloud only), they are not relevant for fog computing. In this paper we consider only combinations that include the layer of fog nodes, which leads to exactly these four categories.

The categories are further subdivided into subcategories. The distinction between the subcategories is based on the purpose of the optimization, that is, the type of decisions resulting from optimization. The subcategories are based on the results of the literature search. It is possible that further work will lead to new subcategories.

As can be seen in Fig. 3, the first two categories have only one subcategory each. Moreover, these two subcategories are strongly related, since both are about offloading compute tasks from the end devices to some other resources, the difference being only if the other resources also include the cloud or only the fog nodes. The third category has five subcategories, while the fourth category has only one subcategory. Moreover, subcategories 3.5 and 4.1 are strongly related, since both are about distributing data and/or applications among some resources, the difference being only if these resources also include the cloud or only the fog nodes.

[2]Only those layers that are actively involved in the optimization are considered. If, for example, data are distributed among the fog nodes to make the data available for access by end devices, then the end devices ultimately use the data, but are not actively involved in the optimization approach.

Table 1: Categories and subcategories of optimization problems in fog computing

Category	**Subcategory**	**Decisions resulting from the optimization**
1. Optimization involving all three architectural layers	1.1. Offloading computing tasks from end devices to fog nodes and cloud	• Whether to offload computing tasks from the end devices to the fog nodes • Whether to further offload computing tasks from the fog nodes to the cloud
2. Optimization involving end devices and fog nodes	2.1. Offloading computing tasks from end devices to fog nodes	• Whether to offload computing tasks from the end devices to fog nodes • To which fog nodes the tasks should be offloaded
3. Optimization involving fog nodes only	3.1. Scheduling within a fog node	• How to prioritise incoming computing tasks within a fog node • How to assign tasks to computing resources within a fog node
	3.2. Clustering of fog nodes	• How to determine the size of a cluster of fog nodes to handle the relevant requests
	3.3. Migration between fog nodes	• Whether to migrate data / applications between fog nodes or to leave them where they are
	3.4. Distributing physical resources (prior to operation)	• Where to place physical resources in the network • With which computing resources should fog nodes be equipped
	3.5. Distributing applications / data among fog nodes	• Which fog node should host which applications and which data
4. Optimization involving fog nodes and cloud	4.1. Distributing applications / data among fog nodes and cloud	• Whether to place data / applications on the individual fog nodes or in the cloud

Table 1 summarizes the categories (first column) and subcategories (second column). The third column describes the decisions resulting from the optimization approaches of the given subcategory, which forms the basis for the classification into subcategories.

For all subcategories – with one exception –, optimization is performed *during operation* of the relevant cloud and/or fog services. The exception is subcategory 3.4 (“Distributing physical resources”), which takes place *prior to the operation* of the service, i.e., when the service is designed or deployed.

5. Metrics

When looking at the publications, it quickly becomes clear that authors use different terms to describe the same concepts, but also use the same term to describe different concepts. This applies in particular to the metrics used in fitness functions and constraints. Therefore, uniform terms had to be defined for these metrics to foster comparability of the publications. The representation in different degrees of detail is an additional problem. For example, some papers distinguish between CPU, RAM and storage when describing the resources of fog nodes, others use abstract resource types (e.g., $R_1, R_2, \ldots$) or combine the description of individual resources into a single value.

Tables 2-3 give an overview of the metrics that we identified from our literature review. This includes metrics for time (latency), energy consumption, profit and cost, characteristics of the devices, and other miscellaneous metrics. As we will see in Sec. 6, all the used objective functions and most of the used constraints are based on these metrics. There is a small number of constraints that are not related to these metrics, which will be shown in Section 6.5.

6. Literature mapping

In this section, we show the mapping of the reviewed papers on the taxonomy introduced in Section 4 and the metrics introduced in Section 5. In particular, Subsections 6.1-6.4 correspond to the four categories of the taxonomy, giving more details about the category and its subcategories, highlighting further variations within some subcategories, and showing the metrics that papers in those subcategories use. Subsection 6.5 provides information about the use of constraints not related to metrics (irrespective of the categories of the taxonomy).

Table 2: Metrics

ID	Description
Latency-related metrics	
Time$_{\text{ED-FN}}$	Time of data transfer between end device and fog node
Time$_{\text{FN}}$	Time of executing a task in the fog node
Time$_{\text{ED}}$	Time of executing a task in the end device
Time$_{\text{C}}$	Time of executing a task in the cloud
Time$_{\text{FN-C}}$	Time of data transfer between fog node and cloud
Time$_{\text{mig}}$	Time of migration of applications between fog nodes
Energy-related metrics	
Energy$_{\text{ED-FN}}$	Energy of data transfer between end device and fog node
Energy$_{\text{FN}}$	Energy of executing a task in the fog node
Energy$_{\text{ED}}$	Energy of executing a task in the end device
Energy$_{\text{C}}$	Energy of executing a task in the cloud
Energy$_{\text{FN-C}}$	Energy of data transfer between fog node and cloud
Energy$_{\text{pos}}$	Energy of positioning a mobile fog node
Energy$_{\text{gained}}$	Energy that is gained in the end device
Energy$_{\text{FN-FN}}$	Energy of data transfer between fog nodes
Profit- and cost-related metrics	
Profit$_{\text{FN}}$	Profit gained by providing the fog nodes
Profit$_{\text{C}}$	Profit gained by providing cloud services
Cost$_{\text{FN}}$	Costs to operate a fog node
Cost$_{\text{C}}$	Costs to operate a cloud service
Cost$_{\text{ED}}$	Costs of processing tasks in the end device
Cost$_{\text{rej}}$	Costs due to rejected tasks
Cost$_{\text{queue}}$	Costs of tasks in queue
Cost$_{\text{sec}}$	Costs of security measures
Cost$_{\text{mig}}$	Costs of migration of applications between fog nodes

Table 3: Metrics (continued)

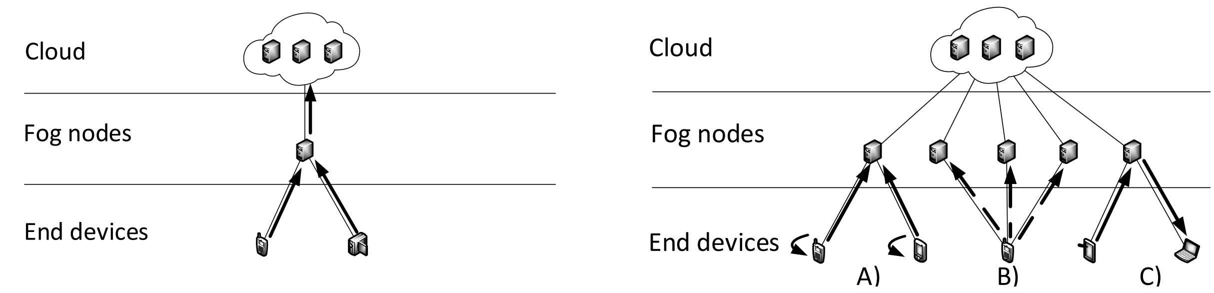
ID	Description
Metrics related to device characteristics	
Load$_{\text{FN}}$	Load of the fog node
Load$_{\text{C}}$	Load of the cloud
Load$_{\text{ED}}$	Load of the end device
ServRate$_{\text{FN}}$	Service rate, tasks executed in the fog node per time unit
ServRate$_{\text{C}}$	Service rate, tasks executed in the cloud per time unit
ServRate$_{\text{ED}}$	Service rate, tasks executed in the end device per time unit
DataRate$_{\text{ED-FN}}$	Bandwidth of the transmission channel between end device and fog node
DataRate$_{\text{FN-C}}$	Bandwidth of the transmission channel between fog node and cloud
Additional metrics	
TaskRatio$_{\text{ED}}$	Percentage of tasks executed on end device
TaskRatio$_{\text{FN}}$	Percentage of tasks executed on fog node
DataTrans$_{\text{net}}$	Total traffic in the entire network
Reliab$_{\text{FN}}$	Measure of reliability of fog node
Quality$_{\text{result}}$	Quality, i.e., deviation from perfect result
ConDur$_{\text{ED-FN}}$	Connection duration between end device and fog node
MaxUsers$_{\text{FN}}$	Maximum number of users for fog node
Count$_{\text{FN}}$	Number of fog nodes that are provided
TaskRatio	Percentage of requests executed by fog nodes
CO2$_{\text{FN}}$	Carbon balance of the fog nodes
Availab$_{\text{FN}}$	Measure of the availability of fog nodes
Security$_{\text{FN}}$	Measure of security provided by fog nodes

6.1. Optimization involving all three architectural layers

Fig. 4a shows schematically the optimization involving all three architectural layers. Tasks from end devices can be offloaded to fog nodes as well as to central cloud services. The decision as to whether the tasks should be passed on to the cloud services is typically made by the fog nodes. Reasons for offloading tasks could be, for example, the limited battery power or processor capacity in the end device. However, it is also possible that the energy or the time required to send the data exceeds the energy or time gained by using a more powerful node, so a local computation in the end device is preferable. This makes offloading decisions non-trivial.

As an example, Chen et al. address the distribution of independent computing tasks of a user over their end device, a Computing Access Point (CAP) representing a fog node, and a remote cloud server [22]. The objective of the optimization approach is to minimize the energy consumption of these three components. To that end, the user's mobile device has to decide first whether the computation is to be executed locally on the device or offloaded to the CAP. The distribution of tasks between the CAP and the cloud server is decided afterwards.

The problem variants formulated in this category differ mainly in the metrics they use in their constraints and objective functions. Table 4 shows the papers in this category and the metrics they use. In this and the next tables, the following symbols depict the role of a metric in a given paper: "→" means that the metric is used in the objective function, "□" means that the metric is used in the constraints, and "⍈" means that the metric is used both in the objective function and in the constraints.



(a) Subcategory 1.1: Offloading to fog nodes and cloud

(b) Subcategory 2.1: Offloading to fog nodes

Figure 4: Offloading of computing tasks from the end devices

Table 4: Metrics used in subcategory 1.1.

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{ED}$	$Time_{C}$	$Time_{FN\text{-}C}$	$Energy_{ED\text{-}FN}$	$Energy_{FN}$	$Energy_{ED}$	$Energy_{C}$	$Energy_{FN\text{-}C}$	$Load_{FN}$	$Load_{C}$	$Load_{ED}$	$DataRate_{ED\text{-}FN}$	$DataRate_{FN\text{-}C}$	$ServRate_{C}$	$ServRate_{FN}$	$ServRate_{ED}$	$Profit_{FN}$	$Cost_{FN}$	$Profit_{C}$	$Cost_{C}$	$Security_{FN}$	$Cost_{rej}$
[23]	□	□	□	□	□	→	→	→	→	→														
[24]											□									→		→		
[25]	□	□		□	□	→				→				□										
[22]	→	→	→	→	→	→	→	→	→	→														
[26]	→	→		→	→																			
[27]																□	□		→		→			
[28]	□	□		□	□																	→		
[29]	→	→	→	→	→						□						□	□						
[30]	→	→	→	→	→						□	□	□											
[31]	→	→	→	→	→	→	→	→	→	→	□			□										
[32]	□	□		□	□	→	→		→	→	□	□		□										
[33]	→	→	→	→	→	→	→	→	→	→														
[34]	□	□		□	□		→		→		□	□			□									
[35]		→		→			⍈		⍈		⍈													
[36]	⍈	⍈		⍈	⍈										□									
[37]	□		□								□								→	→				
[38]	→	→		→	→							□				□	□	□		→		→		
[18]	□	□	□	□	□						□	□	□						→		→			
[39]	→	→	→	→	→	→	→	→	→	→	□			□										
[40]	□	□				⍈					□			□										
[41]											□					→	→							
[42]	→	→	→	→	→						□													
[43]	□	□	□	□	□												□			→		→		
[44]	→				→						□	□												
[45]											□			□						→		→		
[46]											□	□								→		→		
[47]											□									→		→		
[48]	⍈	⍈	⍈	⍈	⍈	□					□			□	□									
[49]	→	→	→	→	→																			
[50]	⍈	⍈	⍈	⍈	⍈	⍈	→	→	→	→	□			□										
[51]						→				→	→	→		□	□									
[52]						→	→		→	→														
[53]	→	→	→	→	→																			
[54]	⍈	⍈		⍈	⍈	⍈	→		→	→	□			□										
[55]	→	→	→	→	→	→	→	→	→	→	□			□	□									
[56]							□							□					→	→	→	→		
[57]		□		□	□		→		→	⍈	□	□			□									
[58]	→	→	→	→	→	→	→	→	→	→														
[59]	→	→	→	→	→	→	→	→	→	→				□	□		□	□		→		→		

Table 4 (continued)

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{ED}$	$Time_{C}$	$Time_{FN\text{-}C}$	$Energy_{ED\text{-}FN}$	$Energy_{FN}$	$Energy_{ED}$	$Energy_{C}$	$Energy_{FN\text{-}C}$	$Load_{FN}$	$Load_{C}$	$Load_{ED}$	$DataRate_{ED\text{-}FN}$	$DataRate_{FN\text{-}C}$	$ServRate_{C}$	$ServRate_{FN}$	$ServRate_{ED}$	$Profit_{FN}$	$Cost_{FN}$	$Profit_{C}$	$Cost_{C}$	$Security_{FN}$	$Cost_{rej}$
[60]	→	→	→	→	→	→	→	→	→	→														
[61]	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$		$\boxed{\rightarrow}$	$\boxed{\rightarrow}$						□	□		□	□									
[62]	□	□		□	□						□	□		□	□					→		→		
[63]	→	→	→	→	→	→	→	→	→	→				□			□							
[64]	→	→	→	→		□		□									□							→
[65]	□	□	□	□	□	$\boxed{\rightarrow}$		$\boxed{\rightarrow}$		$\boxed{\rightarrow}$			□											
[66]	→	→	→	→	→						□			□										
[67]	□	□	□	□	□						→	→	→											
[68]	→	→		→	→		→		→		□	□		□	□									
[69]											□									→			→	
[70]	→	→		→	→																			

As we can see in the table, the focus in this category is on optimizing latency (e.g.: [30]) or energy consumption (e.g.: [23]). In most cases, latency or energy is not only optimized for the end devices, but for the entire continuum from end device to cloud. The constraints mentioned most often express that the load assigned to a fog node must not exceed its capacity (e.g.: [34]).

6.2. Optimization involving end devices and fog nodes

Fig. 4b shows the distribution of the computing load from the end devices between the lower two architectural layers. Several variants can be distinguished within this subcategory, as shown in Fig. 4b. Variant A) deals with the question of whether a computing task should be performed by the end device or by a fog node. Since it can also be useful to only offload some parts of the whole task, the more general question is which parts of the task to offload. Variant B) refers to the selection of the right fog node for a task. Fog nodes can differ, for example, in their resources, utilization, and geographical position, so that the choice of fog node matters. Variants A) and B) can also be combined and treated in one approach. In variant C), fog nodes do not execute tasks themselves but only coordinate between end devices. For this purpose, end devices send their computation tasks to the fog nodes, which in turn distribute the computing load over other end devices. Alternatively, the fog nodes only establish the connection between individual end devices by exchanging metadata with them, and the actual payload data are transferred directly between the end devices.

Since variants A), B), and C) can be combined arbitrarily, they do not represent disjoint sub-subcategories within subcategory 2.1. and thus they are not further subdivided in the taxonomy. The description of the variants serves only to show the scope of subcategory 2.1.

Tao et al. give an example for this category presenting an approach that deals with the distribution of the computing load between end devices and fog nodes [71]. The aim of this approach is to minimize the energy consumption in the execution of computing tasks for a number of mobile devices, taking into account resource constraints and an upper limit on latency. For this purpose, the percentage of each task that is offloaded from the end devices to the fog nodes is determined. This approach is an example for variant A). Mai et al. present an approach that is a representative of variant B) within this category [72]. This approach is about IoT devices generating tasks to be offloaded to the fog nodes. An arbitrary number of fog nodes are available for computation and the aim of optimization is to select one of them so that latency is minimized. The approach of Kattepur et al. is an example of variant C) of this category [73]. This approach is about robots, which represent the end devices, sending computation tasks to a coordinating fog node. The coordinating fog node has an overview of all available computation fog nodes and distributes the tasks over these. When the computation is completed, the coordinating fog node sends the result back to the end devices. The aim of the optimization is to minimize latency and energy consumption.

Table 5: Metrics used in subcategory 2.1.

Paper	$\text{Time}_{\text{ED-FN}}$	Time_{FN}	Time_{ED}	$\text{Energy}_{\text{ED-FN}}$	$\text{Energy}_{\text{FN}}$	$\text{Energy}_{\text{ED}}$	$\text{Energy}_{\text{pos}}$	Load_{FN}	Load_{ED}	$\text{ConDur}_{\text{ED-FN}}$	$\text{DataRate}_{\text{ED-FN}}$	Cost_{FN}	$\text{Quality}_{\text{result}}$	$\text{ServRate}_{\text{ED}}$	$\text{ServRate}_{\text{FN}}$	$\text{Energy}_{\text{gained}}$	$\text{TaskRatio}_{\text{ED}}$	$\text{Profit}_{\text{FN}}$	Cost_{ED}	$\text{Cost}_{\text{queue}}$	Cost_{rej}	$\text{DataTrans}_{\text{net}}$	$\text{Reliab}_{\text{FN}}$
[74]	□	□	□	⍈		→																	
[75]								□										→					
[76]	□	□		⍈																			
[77]	□	□	□	⍈				□															
[78]	□	□	□	⍈		→					□												
[79]	⍈	⍈		□	□																		
[80]								⍈			□				□								
[81]	□	□	□	→		→		□						□	□								
[82]												→							→				
[83]	→	→		□							□												
[84]	→	⍈		□		□					□												
[85]												→								→			
[86]								⍈			□				□								
[87]				□		□						→											
[88]	□	□	□	⍈		→					□												
[89]	→																						
[90]	→	→						→															
[91]								□				→						→					
[92]	→	→	→	□		□																	
[93]	⍈	⍈	⍈	□		□			□												→		
[94]								□						→									
[95]	⍈	⍈	⍈	⍈	⍈	⍈																	
[96]	□	□	□	→	→	→					□												
[97]	□	□		⍈		→					□												
[98]	□	□	□	⍈		→		□															
[99]	□	□	□	→		→		□															
[100]				⍈		→			□														
[101]		□										→											
[73]	→	→	→	→	→	→		□															→
[102]	→		→								□												
[103]		□						□			□	→						→					
[104]	→	→	→																				
[105]	⍈	⍈		□	□																		
[106]	⍈	⍈			□																		
[107]	□	□	□	⍈		→									□								
[108]	→	→	→	→	→	→																	
[109]	□	□		→	→			□															
[110]	□			→		→		□															
[111]	□	□	□	→	→																		
[112]	□	□	□	→		→		□			□												
[113]	→	→	→	→	→	→		□															
[114]			□	⍈	→	□		□															
[115]	□	□	□					□				→											
[116]	⍈	⍈																					
[117]	→	→		⍈																			
[118]	□	□	□	⍈	→	→		□	□														
[119]	⍈	⍈	→								□												
[120]	→	→	→	⍈		→		□			□												
[121]	→	→	→	⍈		→		□															
[122]		→	→	→		→			□														
[123]								□				→											
[124]				⍈	→	→		□															
[125]	⍈	⍈	→					□				→											
[126]	□	□	□																			→	
[71]	□	□	□	⍈	⍈	⍈		□			□												
[127]	⍈	⍈	⍈					□															
[128]	→	→			□																		

Table 5 (continued)

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{ED}$	$Energy_{ED\text{-}FN}$	$Energy_{FN}$	$Energy_{ED}$	$Energy_{pos}$	$Load_{FN}$	$Load_{ED}$	$ConDur_{ED\text{-}FN}$	$DataRate_{ED\text{-}FN}$	$Cost_{FN}$	$Quality_{result}$	$ServRate_{ED}$	$ServRate_{FN}$	$Energy_{gained}$	$TaskRatio_{ED}$	$Profit_{FN}$	$Cost_{ED}$	$Cost_{queue}$	$Cost_{rej}$	$DataTrans_{net}$	$Reliab_{FN}$
[129]								□			□				□			→					
[130]				→	→			⍈													→		
[131]	□	□			⍈						□												
[132]	□	□	□	→		→						□											
[133]											□	→		→				→					
[134]	□													→									
[135]	→	→	→								□												
[136]	→				→			⍈															
[137]	□	□	□	→	→	→		□															
[138]								□														→	
[139]	□	□	□	⍈	→	→		□															
[140]	□	□	□	⍈	→	→		□	□														
[141]	→	→	→								□												
[142]	⍈	⍈	⍈					□															
[143]	□	□	□	□		□		□						→									
[144]	□	□	□	⍈		→		□															
[145]	→	→	→	→	→	→																	
[146]	⍈	⍈	⍈	→	→	→		□															
[147]	□	□	□	→		→		□															
[19]		□		⍈	→			□															
[148]		□	□																				→
[149]				→	⍈	→		□			□												
[150]	→	→						□			□							→					
[151]	□	□	□									→											
[152]	⍈		⍈	□		□		□	□														
[153]	□	□	□	⍈		⍈		□	□							→							
[154]								□			□							→					
[155]		□		⍈				□			□												
[156]												→						→					
[157]		→						□															
[158]	□	□	□					→															
[159]	□	□	□	→		→					□												
[160]	□	□	□	→		→		□			□												
[161]	□	□	□	→		→		□															
[162]	□	□		⍈				□			□												
[163]	⍈	⍈	⍈	⍈	⍈	⍈																	
[164]	□	□	□	→		→		□			□												
[165]	□	□		→	→	→																	
[166]	→	→	→	→	→	→																	
[167]	□	□	□	⍈	→	⍈		□			□												
[168]	⍈	⍈	⍈	⍈		⍈			□														
[169]	⍈	⍈						□					⍈										
[170]	□	□	□	□				→															
[171]	→	→	→	→	→	→																	
[172]	□	□	□	→		→																	
[173]	□	□	□														→	→					
[174]								⍈			□												
[175]	⍈	⍈						□			□												
[176]	□	□	□		→				□		□												
[177]	□	□	□	⍈	→	→		□			□												
[178]	□	□	□	⍈	⍈	⍈																	
[179]			□	⍈	→	□		□															
[180]	□	□	□	⍈	→	⍈																	
[181]	→	→	→	→	→	→		□			□												
[182]	⍈	⍈	⍈	□	□				□														
[183]	□	□						□			⍈	→											
[184]	→	→	→	□	□	□																	

Table 5 (continued)

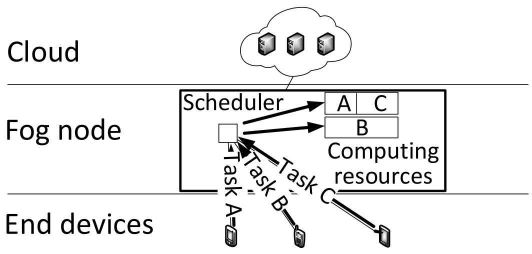
Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{ED}$	$Energy_{ED\text{-}FN}$	$Energy_{FN}$	$Energy_{ED}$	$Energy_{pos}$	$Load_{FN}$	$Load_{ED}$	$ConDur_{ED\text{-}FN}$	$DataRate_{ED\text{-}FN}$	$Cost_{FN}$	$Quality_{result}$	$ServRate_{ED}$	$ServRate_{FN}$	$Energy_{gained}$	$TaskRatio_{ED}$	$Profit_{FN}$	$Cost_{ED}$	$Cost_{queue}$	$Cost_{rej}$	$DataTrans_{net}$	$Reliab_{FN}$
[185]	→	→	→								□												
[186]	→	→						□			□												
[187]	→	→	→					□			□												
[188]	□	□	□	→	→	→		□															
[189]		□										→											
[190]	□	□		→	→			□															
[191]	→	→						□			□												
[192]	→	→	→	→	→	→		□															
[193]				$\boxed{\rightarrow}$			□																
[194]								□				→											
[195]	□	□	□	□	□	□												→					
[196]	→	→																					
[197]	→	→																					→
[198]		□		□	□							→									→		
[199]	□	□		$\boxed{\rightarrow}$				□															
[200]	□	□	□	□		□		□			□						→						
[201]		→		→	→																	→	
[202]	□			$\boxed{\rightarrow}$	□	□																	
[203]	□	□		→	→			□			□												
[72]	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$																					
[204]	→							□			□							→					
[205]	□		□	→		→					□												
[206]	→	→						□															
[207]	→	→	→	□	□	□		□															
[208]	→	→						□			□												
[209]				$\boxed{\rightarrow}$	$\boxed{\rightarrow}$	□	→																
[210]	□	□	□	→	→	→				□													
[211]	□			$\boxed{\rightarrow}$	□	□																	

Table 5 shows the papers in this category and the metrics they use. Similar to the previous category, most approaches are about optimizing metrics related to latency (e.g. [93]) or energy consumption (e.g. [110]). However, in this category the latency and the energy consumption also appear as constraints in the table (e.g.: [163]). This is the case, for example, if the total energy consumption is to be minimized on the one hand, but on the other hand the battery level of the fog nodes or of the end devices restricts the distribution of the computing load. In addition, many authors formulate constraints referring to the load of the fog nodes.

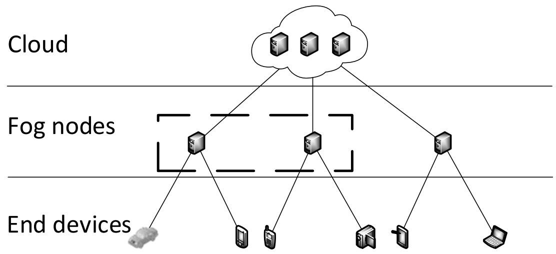
6.3. Optimization involving fog nodes only

Multiple optimization problems can be differentiated that deal with the fog layer. Subcategory 3.1, shown in Fig. 5a, is concerned with the scheduling of requests from end devices within a specific fog node[3]. The aim is to prioritize the tasks and assign them to the available resources within the fog node so that, for instance, the time constraints of all requests are met or the average response time is minimized. As an example, Zhang et al. present an approach for the scheduling of computing tasks within a fog node, with the objective of reducing the operating costs of the fog node while ensuring that all incoming tasks are processed within their respective deadlines [212].

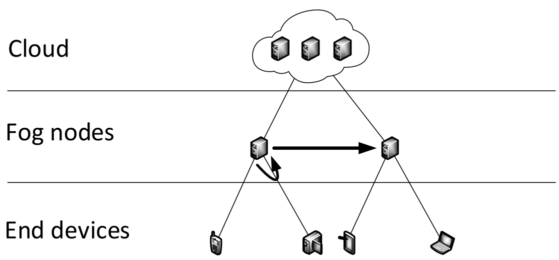
[3] In the literature, the term "scheduling" sometimes refers to the distribution of tasks over multiple fog nodes. Within our taxonomy, these approaches fall under subcategory 2.1 (Offloading of computing tasks from end devices to fog nodes).



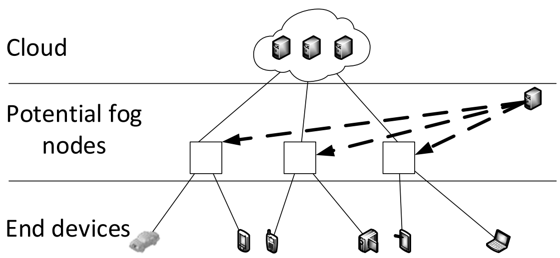
(a) Subcategory 3.1: Scheduling within a fog node



(b) Subcategory 3.2: Clustering of multiple fog nodes



(c) Subcategory 3.3: Deciding on a possible migration



(d) Subcategory 3.4: Distributing physical resources

Figure 5: Subcategories 3.1–3.4 of the category "Optimization involving fog nodes only"

Table 6: Metrics used in subcategory 3.1.

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Energy_{ED\text{-}FN}$	$Energy_{FN}$	$Energy_{ED}$	$Load_{FN}$	$DataRate_{ED\text{-}FN}$	$Energy_{FN\text{-}FN}$	$Profit_{FN}$	$Cost_{FN}$	$Reliab_{FN}$	$ServRate_{FN}$	$Cost_{sec}$
[213]			→	→	→							→	→
[214]	□	□	→	→									
[215]	→	→				□							
[216]	□	□		□								→	
[212]		□				□				→			
[217]							□		→	→			
[218]	□	□				□	□		→				
[219]				→		→							
[220]	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$				□	□						
[221]		□		→				→					
[222]	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$									→		
[223]		→		→						→			
[224]		$\boxed{\rightarrow}$		→						→			
[225]	→	→											

Table 7: Metrics used in subcategory 3.2.

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{C}$	$Time_{FN\text{-}C}$	$Energy_{FN}$	$Load_{FN}$	$DataRate_{ED\text{-}FN}$	$DataRate_{FN\text{-}C}$	$ServRate_{FN}$	$DataTrans_{net}$
[226]	→	→								
[227]					$\boxed{\rightarrow}$	□				
[228]					$\boxed{\rightarrow}$	□				
[229]	→	→	→	→						
[230]						□	□		→	
[231]	→	→								→
[232]	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$		$\boxed{\rightarrow}$		□	□	□		

Table 8: Metrics used in subcategory 3.3.

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Energy_{FN}$	$Load_{FN}$	$DataRate_{ED\text{-}FN}$	Kosten1	$Cost_{FN}$	$DataTrans_{net}$	$MaxUsers_{FN}$	$Time_{mig}$
[233]								→		
[234]				□	□			→		
[20]	→			□						→
[235]	→	→	→	□	□			→		
[236]	→	→		□						
[237]	$\boxed{\rightarrow}$	→			□				→	
[238]	→	→				□				
[239]				□			→			

Table 6 shows the papers in this subcategory and the metrics they use. In this subcategory, there is no metric that would be optimized much more frequently than others. Beside metrics related to latency and energy consumption, the optimization of other metrics also plays an important role. For example, the approach of Wang et al. is about maximizing the profit that can be achieved by operating the fog nodes [218]. To make profit by processing computation tasks, the tasks must be distributed among the computation resources in such a way that their respective deadlines for completion are met.

Fig. 5b shows subcategory 3.2: the clustering of multiple fog nodes. The aim of clustering is to efficiently bundle the storage and computing resources of individual fog nodes so that the resulting cluster can process requests from end devices with sufficiently low latency. The cluster can be formed dynamically, re-calculating the composition of the cluster for each request [227]. In another variant, clusters are retained over a longer period of time. The cluster processes a number of tasks and is only adapted if, for example, fog nodes fail or new ones are added to the network [231]. Oueis et al. describe the dynamic clustering of fog nodes for each task from the end devices [226]. The objective of this approach is to minimize the energy consumption in the cluster, while ensuring that all deadlines for completing the computations are met.

Table 7 shows the papers in this category and the metrics they use. The main focus in this subcategory is on optimizing latency. For example, the approach of Lee et al. aims at optimizing the response time for a request sent by the end devices [229]. One fog node receives a computation request. This fog node selects the neighboring fog nodes and distributes the computation load among them and the cloud. The response time consists of the time for the computation in the fog node and the cloud and the time for data transmission.

Fig. 5c shows subcategory 3.3, which deals with possible migrations of applications or data between fog nodes. This applies, for example, to the case of a mobile end device that is only briefly in the vicinity of a fog node, leaves it again, and approaches another fog node. The optimization problem is about deciding whether the data or applications should remain on a fog node or be migrated to another, following the end device. For instance, the approach of Yao et al. deals with the question whether a virtual machine (VM) should be left on a fog node or migrated [234]. A VM is assigned to a moving vehicle and can run applications for this vehicle in order to relieve its resources. The objective is to minimize network costs, consisting of execution costs and migration costs. Execution costs include the communication of the vehicle with the VM assigned to it. If the VM remains on a fog node, the execution costs increase as the distance to the vehicle increases. On the other hand, if the VM is migrated to "follow" the vehicle, this incurs migration costs.

Table 8 shows the papers in this subcategory and the metrics they use. Latency also plays an important role in this category. For example, Sun et al. describe in their optimization approach a trade-off between the duration for data transfer between the end device and the fog node and the time for a migration of data from one fog node to another [20]. The migration of the data must take into account the capacities of the fog nodes which limit the load that may be assigned to a fog node. In addition to latency, the minimization of all traffic in the network is also an important concern in this category [233, 235].

Subcategory 3.4 deals with the distribution of physical resources among the fog nodes prior to operation (Fig. 5d). An example is the question of where in a network an additional fog node should be placed to relieve the existing fog nodes and minimize the latency with which end devices access the fog nodes. Another example is the question of what capacity the individual fog nodes require to be able to process the requests with acceptable performance. Xu et al. consider a Wireless Metropolitan Area Network (WMAN) consisting of hundreds or thousands of access points (APs) [240]. These access points provide wireless network access for mobile devices and can be equipped with cloudlets that act as fog nodes. The approach aims at placing such cloudlets at strategic locations in the network, leading to optimal resource utilization and a reduction of the average latency with which users access cloudlets. The cloudlets, which differ in their capacities, are sorted according to their capacities and distributed among the potential locations in an iterative process.

Table 9: Metrics used in subcategory 3.4.

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{ED}$	$Load_{FN}$	$DataRate_{ED\text{-}FN}$	TaskRatio	$Count_{FN}$	$Cost_{FN}$
[240]	→			□				
[241]	→			□	□			→
[242]	→			□				
[243]	→			□	□			→
[244]		→		□				
[245]	→							
[246]					□			→
[247]	□	□		□				→
[248]				□		□	→	→
[249]	□			□			→	
[250]							→	
[251]	→				□			
[252]							→	
[253]	→	→	→	□				
[254]	□	□					→	

Table 10: Metrics used in subcategory 3.5.

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{ED}$	$Time_{FN\text{-}C}$	$Energy_{ED\text{-}FN}$	$Energy_{FN}$	$Load_{FN}$	$DataRate_{ED\text{-}FN}$	$Cost_{FN}$	$ServRate_{FN}$	$ServRate_{ED}$	$Profit_{FN}$	$Profit_{C}$	$Cost_{C}$	$Availab_{FN}$
[255]									→						
[256]									→						
[257]	□	□					□	□	→						
[258]	⍈	⍈	⍈				□			□	□				
[259]	⍈	⍈					□								→
[260]							□		→						
[261]											→				
[262]	□						□	□	→						
[263]	→														
[264]							→								
[265]							□	□			→				
[266]	→							□							
[267]				□			□				→				
[268]	→								→			→	→	→	
[269]					→	→	□								
[270]					→	→	□								
[271]	□						□	□	→						

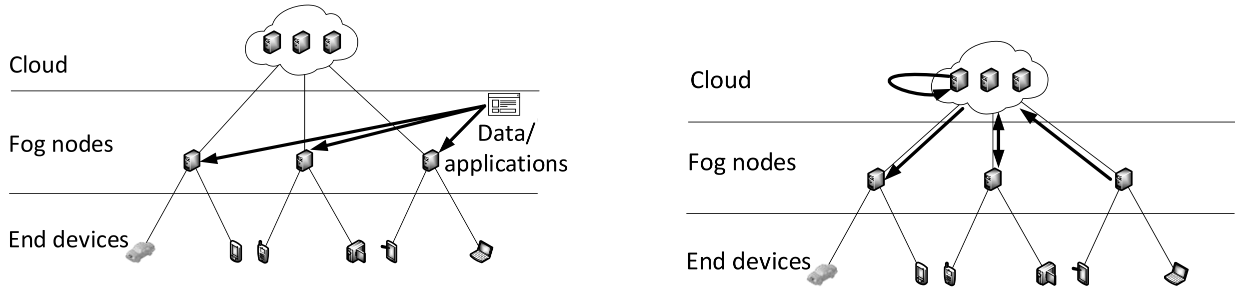
Table 9 shows the papers in this subcategory and the metrics they use. In addition to latency [240], this subcategory mainly addresses the optimization of the number of fog nodes and the costs for operating fog nodes. Premsankar et al. address the last two points [248]. On the one hand this approach is about minimizing the number of deployed fog nodes, on the other hand the operating costs should be minimized. The operating costs consist of the power level required to operate the fog node. In addition, there are the costs that arise when a fog node covers more than one cell and data has to be transferred between the cells. Also in this approach it is emphasized that the capacities of the fog nodes have to be taken into account.

Subcategory 3.5 deals with the distribution of data or applications among the fog nodes (Fig. 6a), so that end devices can access the applications or data via nearby fog nodes. Optimization involves in this subcategory, for example, distributing content among the fog nodes in such a way that the expected latency with which the content will be accessed by end devices is minimized. Gu et al. present an approach for the distribution of VMs of a medical cyber-physical system over a set of base stations [257]. The base stations, which are the fog nodes in this case, differ in the number of connections provided to the end devices and the deployment costs of a VM. The objective is to minimize the costs associated with deploying VMs to base stations and with inter-base-station communication, while ensuring that the resource requirements of the VMs are met.

Table 10 shows the papers in this subcategory and the metrics they use. The metrics to be optimized are quite varied. Beside latency (e.g.: [266]) and energy consumption (e.g.: [269]), also the optimization of costs for operating a fog node is important (e.g.: [268]). The most frequently mentioned constraint relates to the load on the fog nodes, which must not exceed the storage and computing capacities.

6.4. Optimization involving fog nodes and cloud

The only subcategory of this category is shown in Fig. 6b and deals with the distribution of applications or data among the fog nodes and the cloud. The distribution of data or applications can be directed from the cloud to the fog nodes (e.g., distributing content from the cloud to the fog nodes) or vice versa (e.g., migrating resource-hungry application modules from the fog nodes to the cloud), or undirected (e.g., deploying the modules of an application on a combination of fog nodes and cloud services). Skarlat et al. address the placement of services of an IoT application on the fog nodes and the cloud [272, 273]. The objective is to maximize the number of services assigned to the fog nodes, thus minimizing the services assigned to the cloud. The requirements of the individual services with regard to CPU, RAM and storage must be taken into account as well as the deadline for deploying and executing the application. The authors emphasize the advantage of lower latency when using the fog nodes. However, services must be placed in the



(a) Subcategory 3.5: Distribution among fog nodes

(b) Subcategory 4.1: Distribution among fog nodes and cloud

Figure 6: Distributing data or applications

cloud if their processing in the fog nodes is not possible. The individual services of an application can be distributed among a combination of the fog nodes and the cloud. In addition, the allocation of a service can change from time slot to time slot, moving it from the fog nodes to the cloud and vice versa. For this reason, this approach is representative of all three variants.

However, for other approaches in this category, it is clear which of the three variants they belong to. For content caching, it is obvious that content is distributed from the cloud among the fog nodes. For example, the approach of Hou et al. is about reducing the cost of data transfer in the fog nodes. The cost depends on whether data is provided directly by the fog node that receives the request from an end device, the data is provided by a neighboring fog node, or the data needs to be retrieved from the cloud [274, 275]. In order to decide which content will be cached on the fog node, a content popularity estimation is performed based on historical data for similar content.

Table 11 shows the papers in this subcategory and the metrics they use. In addition to latency optimization (e.g. [276]), cost optimization plays an important role in this category. For example, Wang et al. address the minimization of costs [277]. On the one hand, this involves the operating costs for a fog node, which consist of costs for task computation and data transmission, and on the other hand the costs for the migration of services between the fog nodes. Services can be placed both in the cloud and in the fog nodes. Almost all approaches name the capacities of the fog nodes as limitations. One reason for this is that many approaches in the subcategory describe the caching of data or services in the fog nodes. The storage resources of the fog nodes play an important role in deciding whether certain data or services can be cached in the fog nodes.

Table 11: Metrics used in subcategory 4.1.

Paper	$\text{Time}_{\text{ED-FN}}$	Time_{FN}	Time_{ED}	Time_{C}	$\text{Time}_{\text{FN-C}}$	$\text{Energy}_{\text{ED-FN}}$	$\text{Energy}_{\text{FN}}$	Load_{FN}	$\text{DataRate}_{\text{ED-FN}}$	$\text{DataRate}_{\text{FN-C}}$	Cost_{FN}	$\text{ServRate}_{\text{ED}}$	CO2_{FN}	$\text{TaskRatio}_{\text{FN}}$	$\text{Profit}_{\text{FN}}$	Load_{C}	Cost_{mig}	TaskRatio
[278]								□					→		→			
[279]	→							→								→		
[273]	□	□	□	□	□			□						→				
[280]								□			→							→
[281]											$\boxed{\rightarrow}$				→			
[21]	□	□						□	□		→				→			
[277]											→						→	
[282]								□			→							
[283]	□	□						□										→
[272]	□	□	□	□	□			□						→				
[274]								□			→							
[284]	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$			□						→				
[285]								□							→			
[286]	→							□										
[287]	$\boxed{\rightarrow}$	$\boxed{\rightarrow}$		$\boxed{\rightarrow}$	$\boxed{\rightarrow}$			□	□	□								

Table 11 (continued)

Paper	$Time_{ED\text{-}FN}$	$Time_{FN}$	$Time_{ED}$	$Time_{C}$	$Time_{FN\text{-}C}$	$Energy_{ED\text{-}FN}$	$Energy_{FN}$	$Load_{FN}$	$DataRate_{ED\text{-}FN}$	$DataRate_{FN\text{-}C}$	$Cost_{FN}$	$ServRate_{ED}$	$CO2_{FN}$	$TaskRatio_{FN}$	$Profit_{FN}$	$Load_{C}$	$Cost_{mig}$	TaskRatio
[288]	→							□										
[289]	→						□	□										
[290]		$\boxed{\rightarrow}$					□	□										
[291]	→				→			□										
[292]						→	→	□										
[293]									□			→						
[294]															→			
[276]	→	→		→	→			□										
[275]								□			→							
[295]	→				→			□										
[296]								□				→						
[297]	□	□	□			□		□	□		→				→			

6.5. Further constraints

Constraints on specific metrics have already been covered in Sections 6.1-6.4. Several papers use further equations and inequalities to ensure some property that solutions must obviously possess but which is not guaranteed by the used solution encoding. Examples of such properties include:

- The distribution of a computing task among several layers must add up to 100 percent
- The capacity provided by a fog node must not be negative
- All incoming requests must be processed by the fog nodes[4]

We omit a comprehensive coverage of the usage of such constraints in the literature because of the limited insights that this would yield. Instead, Table 12 provides an overview of some of the less obvious constraints used in the literature that are not related to metrics.

Table 12: Constraints not related to metrics

Constraint	**Papers**
End device is assigned to exactly one fog node	[54], [165], [178], [224], [238], [234], [269], [270]
Potential fog nodes for executing application	[258], [267], [271], [289], [276]
Potential locations for fog nodes	[240], [242], [253]
Service is placed only once	[260], [265], [281]
Privacy conflicts of two datasets	[136]
Tasks must not be interrupted	[219]

7. Discussion

This section presents the results of a quantitative analysis of the literature mapping, followed by a discussion on the threats to validity.

7.1. Analysis

Fig. 7 provides an overview of the evolution of the number of relevant papers over time. It should be noted that the figure for 2018 does not refer to the complete year, but only until November 1, 2018. The first two relevant approaches

[4] In some approaches, this is relaxed to the minimization of the metric $Cost_{rej}$ (Costs due to rejected tasks)

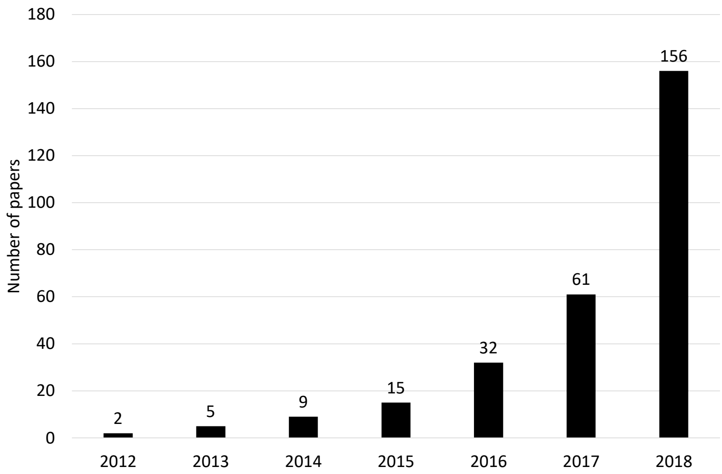


Figure 7: Evolution of the number of relevant papers over time

were presented in 2012. In the following three years the number of papers per year increased to 15 in 2015. The first two conferences specifically dealing with fog computing were held for the first time in 2016, which led to an increase of over 100 percent over the previous year. In 2017, the number of publications again nearly doubled. The number for the first ten months of 2018 shows again a very strong growth and thus underlines the growing importance of optimization approaches in fog computing.

Fig. 8 shows the number of publications in the eight subcategories of the taxonomy. Comparing those numbers, the predominance of subcategories 1.1 and 2.1 is noticeable. These (sub-)categories together account for approx. 68% of the total number of papers. Subcategory 2.1 alone has more representatives than all other subcategories together. The individual subcategories of Category 3 have the smallest number of representatives, which is partly due to the large variance among these approaches, reflected by the subdivision into five subcategories.

Fig. 9 shows which metrics are optimized most frequently by the approaches of the literature. In many cases the authors address the optimization of several metrics; thus, the sum of the values is higher than the number of the examined publications. The different degree of detail with which the metrics are defined should also be noted: e.g., energy consumption is regarded as a separate metric in some approaches, while the cost of the consumed energy is part of a compound cost metric in other approaches. Looking at the values in the figure, it is clear that most authors address the minimization of energy consumption and/or latency. In particular, the focus is on transmission between end devices and fog nodes, as well as on latency and energy consumption during task execution by fog nodes and the end devices. This is due both to the large number of optimization problems relating to the lower two architectural layers and to the criticality of latency and energy consumption in these layers. In addition, the metric of costs for operating fog nodes stands out. The other metrics have relatively low values, which in some cases can be explained by the fact that they are relevant mainly to one specific (sub-)category, as we will analyze next.

Fig. 10 shows the relative frequency with which the approaches of the eight subcategories address the optimization of different metrics. The columns of the diagram refer to the subcategories, the rows to the metrics. The circles located at the intersections of the rows and columns represent with their diameter the relative frequency of the given metric within the subcategory. To keep the figure clear, only those metrics are considered that have a frequency of at least 10 percent in at least one subcategory. Like in the previous figure, the dominance of energy consumption and latency is evident. The metrics that relate to these are the most common in all categories. Energy consumption receives less attention than latency, which may be related to its inclusion in operating costs in some cases. The connection between end devices and fog nodes is particularly frequently dealt with in the literature, while the connection between fog nodes and cloud stands out only in subcategory 1.1. Another important metric is the operating cost of fog nodes, which has a frequency similar to energy consumption and latency in subcategories 3.1, 3.4, and 3.5. The other metrics receive less attention overall and only play a role for individual subcategories. Subcategory 3.5, for example, is particularly aimed at reducing network traffic, while subcategory 3.4 has a focus on the number of fog nodes.

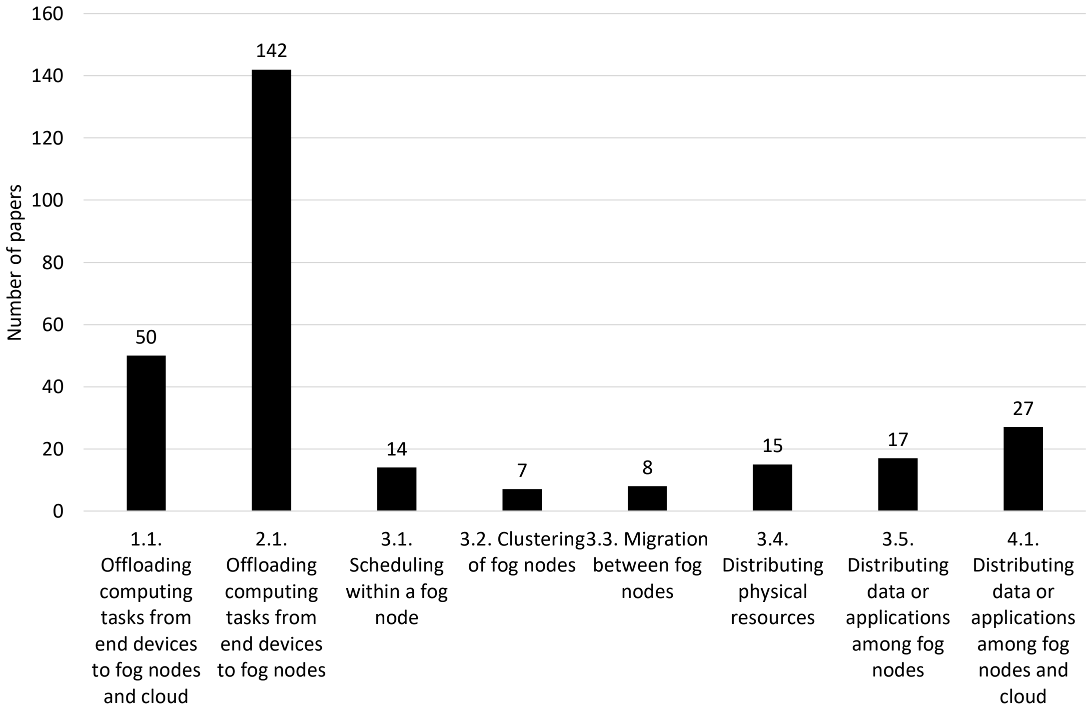


Figure 8: Distribution of the papers across the subcategories of the taxonomy

When looking at the publication venues where the 280 analyzed papers were published, a high variety can be noticed. Specifically, the 280 papers were published in 115 different venues. Table 13 shows the conferences and workshops in which at least 3 papers relevant to this work were published. It is noticeable that the conferences IEEE International Conference on Communications and IEEE Global Communications Conference stand out, with the first having significantly more papers than the second. Table 14 shows the journals in which at least 3 papers relevant to this work were published. In this table the journals IEEE Internet of Things Journal, IEEE Access and IEEE Transactions on Vehicular Technology stand out. In both tables it is noticeable that the number of published papers behind these outstanding venues drops rapidly. A further 90 venues have each published only 1 or 2 relevant papers. 110 papers were published at these other venues.

7.2. Threats to validity

The *internal validity* of our study is influenced by the categorization of publications, which is shaped by our subjective view on the publications. To minimize bias caused by this subjective view, we have discussed the classification of publications between us to achieve a common result. In addition, we presented the taxonomy and the findings from the categorization of papers multiple times to colleagues to obtain their feedback. The taxonomy was developed and continuously improved in an iterative process to ensure that it reflects an appropriate view of the state of research in this area.

External validity is threatened by the possibility that a class of relevant approaches may not have been found and that this distorts the taxonomy and the resulting findings. To address this issue, we combined different search methods (manual search, keyword search, snowball search). Also, we did not limit ourselves to the keyword “fog computing”, but also considered related terms like “edge computing” or “cloudlet”.

Conclusion validity is influenced by the fact that authors sometimes use different terms for the same metric or the same term for different metrics. They have a different understanding of what a term includes, especially the

[5]The conferences ISPA and IUCC were held together in 2017 and the corresponding proceedings were jointly published

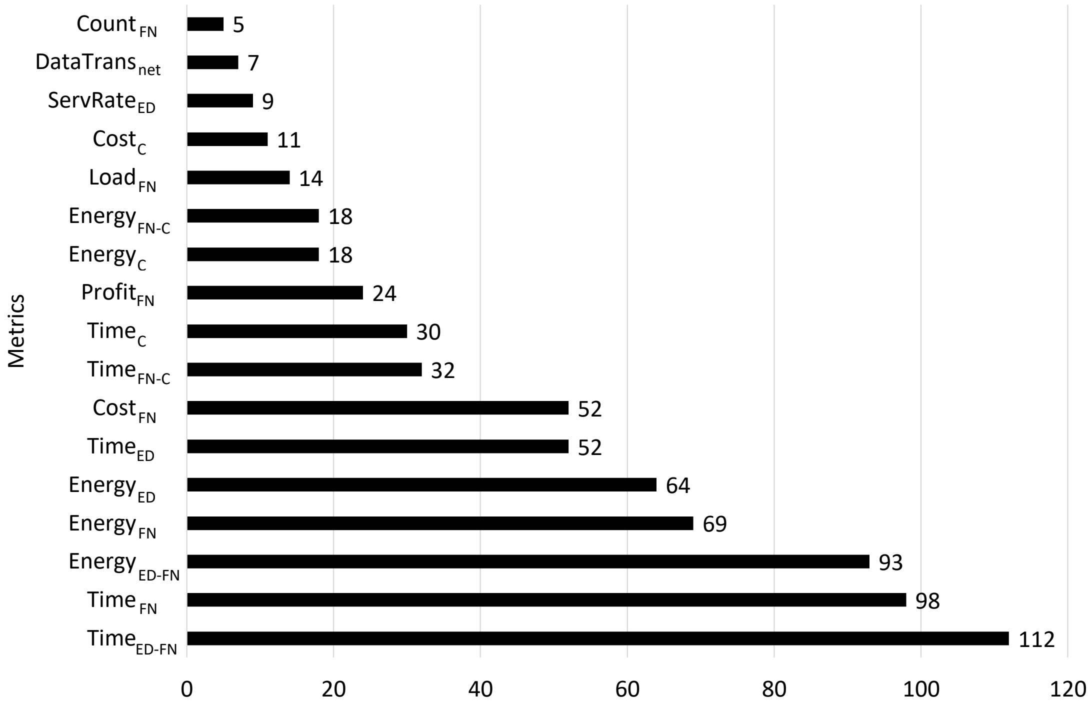


Figure 9: Number of approaches optimizing the metrics. Only metrics considered in at least 5 papers are shown

terms latency and energy consumption are used in many different variants. We addressed this point by defining and consistently using a set of metrics. We use the terms we defined for the metrics instead of the individual terms used in the various publications, thus fostering comparability.

Construct validity is threatened by the "arbitrary" definition of the used taxonomy. To minimize the impact of this threat, we derived the categories of the taxonomy from the logically possible combinations of the involved architectural layers.

8. Conclusions

We have presented the results of a systematic literature review on optimization problems in fog computing. Through a combination of manual search, keyword search and snowball search, we identified 280 relevant publications. Parallel to reviewing these papers, we constructed a taxonomy of optimization problems in fog computing, consisting of 4 categories with altogether 8 subcategories. In addition, we identified 43 different metrics used by the surveyed optimization approaches. We categorized the 280 papers according to the taxonomy of optimization problems and the metrics used in the constraints and objective functions.

The overall findings from our survey can be summarized as follows:

- The number of publications on optimization problems in fog computing has been growing exponentially over the years, underlining the importance of the topic.
- Obtaining an overview of this field of research is challenging for several reasons. Relevant research results are fragmented over different research communities, scattered over a wide variety of journals and conferences. Also the terminology used in the different research communities is different (e.g., fog computing versus edge computing). In particular, there is no generally accepted set of metrics with consistent semantics.
- The most frequently addressed categories of optimization problems in fog computing relate to the offloading of computing tasks from end devices to fog nodes (and potentially also to the cloud). Optimization approaches

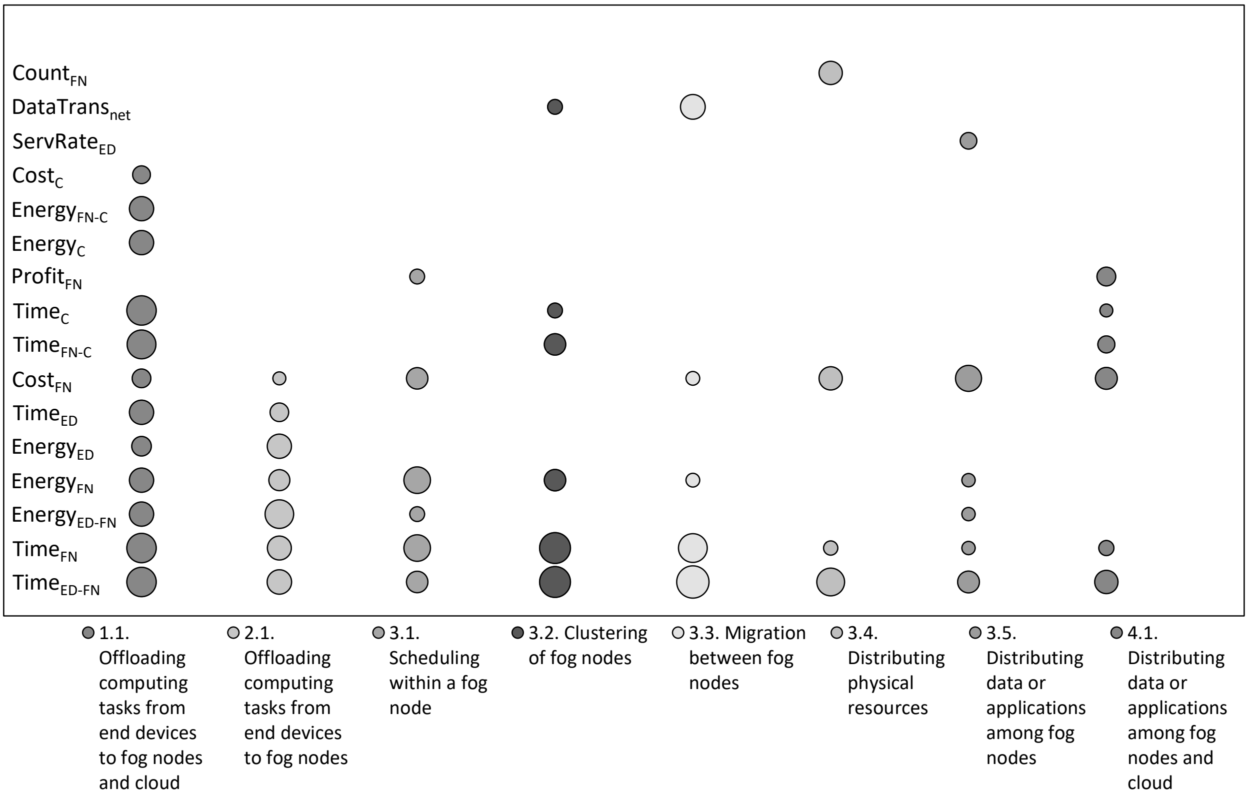


Figure 10: Relative frequency of metrics within each subcategory

that only deal with fog nodes are less numerous but more varied.

- The most widely used metrics are related to timing and energy of end devices, fog nodes, and data transfer among them. In specific categories of optimization problems, further metrics (costs, profit, number of nodes etc.) also play an important role.
- Most optimization approaches described in the literature apply to optimization online during system operation, but there are also approaches for optimizing a fog computing system offline, before its operation starts.

By structuring the state of research in optimization in fog computing according to problem variants and the used metrics, our study made a contribution towards the maturation of this important field. However, several important research directions remain. In particular, a next step can be to survey the different algorithms that solve a particular optimization problem, or family of similar optimization problems in fog computing. Moreover, algorithms solving the same problem (also meaning that they address the same metrics), can be directly compared to each other, for example through appropriate experiments and statistical assessment of the results, to find out what the best algorithms for particular problem variants are. For this purpose, also benchmark problems or problem generators as well as a widely accepted benchmarking methodology would be needed.

Acknowledgement

This work was partially supported by the European Union's Horizon 2020 research and innovation programme under grant 731678 (RestAssured). Useful feedback from colleagues at the Department of Software Systems Engineering, University of Duisburg-Essen, is gratefully acknowledged.

Table 13: Conferences and workshops with the highest number of relevant papers

Venue	**Number of papers**
IEEE International Conference on Communications (ICC)	26
IEEE Global Communications Conference (GLOBECOM)	11
IEEE International Conference on Communications Workshops (ICC Workshops)	7
IEEE Vehicular Technology Conference (VTC)	6
IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)	5
IEEE Wireless Communications and Networking Conference (WCNC)	5
IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)	5
IEEE INFOCOM – IEEE Conference on Computer Communications Workshops (INFOCOM Workshops)	4
IEEE Conference on Computer Communications (INFOCOM)	4
IEEE International Conference on Fog and Edge Computing (ICFEC)	3
IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA) and IEEE International Conference on Ubiquitous Computing and Communications[5] (IUCC)	3
International Conference on Advanced Infocomm Technology (ICAIT)	3

Table 14: Journals with the highest number of relevant papers

Venue	**Number of papers**
IEEE Internet of Things Journal	16
IEEE Access	14
IEEE Transactions on Vehicular Technology	13
IEEE Transactions on Wireless Communications	8
IEEE Journal on Selected Areas in Communications	7
IEEE Transactions on Industrial Informatics	6
IEEE Transactions on Mobile Computing	6
IEEE Transactions on Communications	4
Sensors	4
China Communications	3
IEEE Transactions on Network and Service Management	3
IEEE Transactions on Parallel and Distributed Systems	3
Journal of Network and Computer Applications	3

References

[1] F. Bonomi, R. A. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things, in: MCC@SIGCOMM 2012, 2012, pp. 13–16.

[2] M. Iorga, L. Feldman, R. Barton, M. Martin, N. Goren, C. Mahmoudi, Fog computing conceptual model, recommendations of the national institute of standards and technology, NIST Special Publication (2018) 500–325.

[3] L. M. V. González, L. Rodero-Merino, Finding your way in the fog: Towards a comprehensive definition of fog computing, Computer Communication Review 44 (5) (2014) 27–32.

[4] Z. Á. Mann, Resource optimization across the cloud stack, IEEE Transactions on Parallel and Distributed Systems 29 (1) (2018) 169–182.

[5] Z. Á. Mann, Optimization problems in fog and edge computing, in: R. Buyya, S. N. Srirama (Eds.), Fog and Edge Computing: Principles and Paradigms, Wiley, 2019, pp. 103–121.

[6] R. Mahmud, R. Kotagiri, R. Buyya, Fog computing: A taxonomy, survey and future directions, in: Internet of everything, Springer, 2018, pp. 103–130.

[7] S. Yi, C. Li, Q. Li, A survey of fog computing: Concepts, applications and issues, in: Mobidata@MobiHoc 2015, 2015, pp. 37–42.

[8] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, P. A. Polakos, A comprehensive survey on fog computing: State-of-the-art and research challenges, IEEE Communications Surveys & Tutorials 20 (1) (2017) 416–464.

[9] P. Hu, S. Dhelim, H. Ning, T. Qiu, Survey on fog computing: architecture, key technologies, applications and open issues, Journal of network and computer applications 98 (2017) 27–42.

[10] M. Mukherjee, L. Shu, D. Wang, Survey of fog computing: Fundamental, network applications, and research challenges, IEEE Communications Surveys & Tutorials 20 (3) (2018) 1826–1857.

[11] S. Yi, Z. Qin, Q. Li, Security and privacy issues of fog computing: A survey, in: WASA 2015, Springer, 2015, pp. 685–695.

[12] R. Roman, J. Lopez, M. Mambo, Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges, Future Generation Computer Systems 78 (2018) 680–698.

[13] I. Stojmenovic, S. Wen, X. Huang, H. Luan, An overview of fog computing and its security issues, Concurrency and Computation: Practice and Experience 28 (10) (2016) 2991–3005.
[14] M. Aazam, S. Zeadally, K. A. Harras, Offloading in fog computing for IoT: Review, enabling technologies, and research opportunities, Future Generation Computer Systems 87 (2018) 278–289.
[15] D. Xu, Y. Li, X. Chen, J. Li, P. Hui, S. Chen, J. Crowcroft, A survey of opportunistic offloading, IEEE Communications Surveys & Tutorials 20 (3) (2018) 2198–2236.
[16] Z. Á. Mann, A. Metzger, J. Prade, R. Seidl, Optimized application deployment in the fog, in: 17th International Conference on Service-Oriented Computing, 2019, pp. 283–298.
[17] Z. Á. Mann, Optimization in computer engineering – Theory and applications, Scientific Research Publishing, Incorporated, 2011.
[18] J. Fan, X. Wei, T. Wang, T. Lan, S. Subramaniam, Deadline-aware task scheduling in a tiered IoT infrastructure, in: GLOBECOM 2017, 2017, pp. 1–7.
[19] X. He, Y. Chen, K. K. Chai, Delay-aware energy efficient computation offloading for energy harvesting enabled fog radio access networks, in: VTC Spring 2018, 2018, pp. 1–6.
[20] X. Sun, N. Ansari, PRIMAL: profit maximization avatar placement for mobile edge computing, in: ICC 2016, 2016, pp. 1–6.
[21] W. Tärneberg, A. Mehta, E. Wadbro, J. Tordsson, J. Eker, M. Kihl, E. Elmroth, Dynamic application placement in the mobile cloud network, Future Generation Comp. Syst. 70 (2017) 163–177.
[22] M. Chen, B. Liang, M. Dong, A semidefinite relaxation approach to mobile cloud offloading with computing access point, in: SPAWC 2015, 2015, pp. 186–190.
[23] A. Mtibaa, A. Fahim, K. A. Harras, M. H. Ammar, Towards resource sharing in mobile device clouds: power balancing across mobile devices, Computer Communication Review 43 (4) (2013) 51–56.
[24] F. Chi, X. Wang, W. Cai, V. C. M. Leung, Ad hoc cloudlet based cooperative cloud gaming, in: CloudCom 2014, 2014, pp. 190–197.
[25] Q. Xia, W. Liang, Z. Xu, B. B. Zhou, Online algorithms for location-aware task offloading in two-tiered mobile cloud environments, in: UCC 2014, 2014, pp. 109–116.
[26] X. He, Z. Ren, C. Shi, J. Fang, A novel load balancing strategy of software-defined cloud/fog networking in the internet of vehicles, China Communications 13 (Supplement2) (2016) 140–149.
[27] Y. Liu, M. J. Lee, Y. Zheng, Adaptive multi-resource allocation for cloudlet-based mobile cloud computing system, IEEE Trans. Mob. Comput. 15 (10) (2016) 2398–2410.
[28] Y. Nan, W. Li, W. Bao, F. C. Delicato, P. F. Pires, A. Y. Zomaya, Cost-effective processing for delay-sensitive applications in cloud of things systems, in: NCA 2016, 2016, pp. 162–169.
[29] W. Fan, Y. Liu, B. Tang, F. Wu, H. Zhang, Exploiting joint computation offloading and data caching to enhance mobile terminal performance, in: 2016 IEEE Globecom Workshops, 2016, pp. 1–6.
[30] V. B. C. Souza, W. Ramírez, X. Masip-Bruin, E. Marín-Tordera, G. Ren, G. Tashakor, Handling service allocation in combined fog-cloud scenarios, in: ICC 2016, 2016, pp. 1–5.
[31] M. Chen, M. Dong, B. Liang, Joint offloading decision and resource allocation for mobile cloud with computing access point, in: ICASSP 2016, 2016, pp. 3516–3520.
[32] K. Liang, L. Zhao, X. Zhao, Y. Wang, S. Ou, Joint resource allocation and coordinated com-putation offloading for fog radio access networks, China Communications 13 (2z) (2016) 131–139.
[33] M. Chen, M. Dong, B. Liang, Multi-user mobile cloud offloading game with computing access point, in: Cloudnet 2016, 2016, pp. 64–69.
[34] R. Deng, R. Lu, C. Lai, T. H. Luan, H. Liang, Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption, IEEE Internet of Things Journal 3 (6) (2016) 1171–1181.
[35] V. B. C. Souza, X. Masip-Bruin, E. Marín-Tordera, W. Ramírez, S. Sánchez-López, Towards distributed service allocation in fog-to-cloud (F2C) scenarios, in: GLOBECOM 2016, 2016, pp. 1–6.
[36] S. Rashidi, S. Sharifian, A hybrid heuristic queue based algorithm for task assignment in mobile cloud, Future Generation Comp. Syst. 68 (2017) 331–345.
[37] H. Zhang, Y. Xiao, S. Bu, D. Niyato, F. R. Yu, Z. Han, Computing resource allocation in three-tier IoT fog networks: A joint optimization approach combining stackelberg game and matching, IEEE Internet of Things Journal 4 (5) (2017) 1204–1215.
[38] X. Ma, S. Zhang, W. Li, P. Zhang, C. Lin, X. Shen, Cost-efficient workload scheduling in cloud assisted mobile edge computing, in: IWQoS 2017, 2017, pp. 1–10.
[39] M. Chen, B. Liang, M. Dong, Joint offloading and resource allocation for computation and communication in mobile cloud with computing access point, in: INFOCOM 2017, 2017, pp. 1–9.
[40] A. Al-Shuwaili, O. Simeone, A. Bagheri, G. Scutari, Joint uplink/downlink optimization for backhaul-limited mobile cloud computing with user scheduling, IEEE Trans. Signal and Information Processing over Networks 3 (4) (2017) 787–802.
[41] T. Zhao, S. Zhou, X. Guo, Z. Niu, Tasks scheduling and resource allocation in heterogeneous cloud for delay-bounded mobile edge computing, in: ICC 2017, 2017, pp. 1–7.
[42] D. S. Roy, R. K. Behera, K. H. K. Reddy, R. Buyya, A context-aware, fog enabled scheme for real-time, cross-vertical IoT applications, IEEE Internet of Things Journal.
[43] Y. Nan, W. Li, W. Bao, F. C. Delicato, P. F. Pires, A. Y. Zomaya, A dynamic tradeoff data processing framework for delay-sensitive applications in cloud of things systems, J. Parallel Distrib. Comput. 112 (2018) 53–66.
[44] M. I. Naas, L. Lemarchand, J. Boukhobza, P. R. Parvedy, A graph partitioning-based heuristic for runtime IoT data placement strategies in a fog infrastructure, in: SAC 2018, 2018, pp. 767–774.
[45] M. A. Sharkh, M. Kalil, A quest for optimizing the data processing decision for cloud-fog hybrid environments, in: ICC Workshops 2018, 2018, pp. 1–6.
[46] N. Téllez, M. Jimeno, A. Salazar, E. D. Nino-Ruiz, A tabu search method for load balancing in fog computing, Int. J. Artif. Intell 16 (2).
[47] F. Chi, X. Wang, W. Cai, V. C. M. Leung, Ad-hoc cloudlet based cooperative cloud gaming, IEEE Trans. Cloud Computing 6 (3) (2018) 625–639.

[48] J. Wang, D. Li, Adaptive computing optimization in software-defined network-based industrial internet of things with fog computing, Sensors 18 (8) (2018) 2509.

[49] T. Chen, G. B. Giannakis, Bandit convex optimization for scalable and dynamic IoT management, IEEE Internet of Things Journal.

[50] J. Du, L. Zhao, J. Feng, X. Chu, Computation offloading and resource allocation in mixed fog/cloud computing systems with min-max fairness guarantee, IEEE Trans. Communications 66 (4) (2018) 1594–1608.

[51] K. Kaur, S. Garg, G. S. Aujla, N. Kumar, J. J. P. C. Rodrigues, M. Guizani, Edge computing in the industrial internet of things environment: Software-defined-networks-based edge-cloud interplay, IEEE Communications Magazine 56 (2) (2018) 44–51.

[52] C. Wang, J. Kuo, D. Yang, W. Chen, Green software-defined internet of things for big data processing in mobile edge networks, in: ICC 2018, 2018, pp. 1–7.

[53] T. Chen, Q. Ling, Y. Shen, G. B. Giannakis, Heterogeneous online learning for thing-adaptive fog computing in IoT, IEEE Internet of Things Journal.

[54] Y. Liu, F. R. Yu, X. Li, H. Ji, V. C. M. Leung, Hybrid computation offloading in fog and cloud networks with non-orthogonal multiple access, in: INFOCOM Workshops 2018, 2018, pp. 154–159.

[55] L. Liu, X. Guo, Z. Chang, T. Ristaniemi, Joint optimization of energy and delay for computation offloading in cloudlet-assisted mobile cloud computing, Wireless Networks (2018) 1–14.

[56] S. Meng, Y. Wang, Z. Miao, K. Sun, Joint optimization of wireless bandwidth and computing resource in cloudlet-based mobile cloud computing environment, Peer-to-Peer Networking and Applications 11 (3) (2018) 462–472.

[57] J. Tan, T. Chang, T. Q. S. Quek, Minimum energy resource allocation in FOG radio access network with fronthaul and latency constraints, in: SPAWC 2018, 2018, pp. 1–5.

[58] K. R. Alasmari, R. C. G. II, M. Alam, Mobile edge offloading using markov decision processes, in: EDGE 2018, 2018, pp. 80–90.

[59] L. Liu, Z. Chang, X. Guo, S. Mao, T. Ristaniemi, Multiobjective optimization for computation offloading in fog computing, IEEE Internet of Things Journal 5 (1) (2018) 283–294.

[60] S. Midya, A. Roy, K. Majumder, S. Phadikar, Multi-objective optimization technique for resource allocation and task scheduling in vehicular cloud architecture: A hybrid adaptive nature inspired approach, Journal of Network and Computer Applications 103 (2018) 58–84.

[61] F. Y. Lin, C. Hsiao, Y. Wen, Y. Wu, Optimization-based resource management strategies for 5G C-RAN slicing capabilities, in: ICUFN 2018, 2018, pp. 346–351.

[62] N. M. Randriamasinoro, K. K. Nguyen, M. Cheriet, Optimized resource allocation in edge-cloud environment, in: SysCon 2018, 2018, pp. 1–8.

[63] M. Chen, M. Dong, B. Liang, Resource sharing of a computing access point for multi-user mobile cloud offloading with delay constraints, IEEE Trans. Mob. Comput. 17 (12) (2018) 2868–2881.

[64] L. Liu, Z. Chang, X. Guo, Socially aware dynamic computation offloading scheme for fog computing system with energy harvesting devices, IEEE Internet of Things Journal 5 (3) (2018) 1869–1879.

[65] T. Zhang, C. F. Chiasserini, P. Giaccone, Tame: An efficient task allocation algorithm for integrated mobile gaming, IEEE Systems Journal.

[66] H. Zhao, Y. Wang, R. Sun, Task proactive caching based computation offloading and resource allocation in mobile-edge computing systems, in: IWCMC 2018, 2018, pp. 232–237.

[67] Y. Lin, Y. Lai, J. Huang, H. Chien, Three-tier capacity and traffic allocation for core, edges, and devices for mobile edge computing, IEEE Trans. Network and Service Management 15 (3) (2018) 923–933.

[68] X. Wang, H. Ni, R. Han, X. Huang, Trade-off between service delay and power consumption in edge-cloud computing, International Journal of Innovative Computing, Information and Control 14 (6) (2018) 2011–2024.

[69] Q. Xu, Z. Su, M. Dai, Trustworthy caching for mobile big data in social networks, in: INFOCOM Workshops 2018, 2018, pp. 808–812.

[70] C. Shi, Z. Ren, K. Yang, C. Chen, H. Zhang, Y. Xiao, X. Hou, Ultra-low latency cloud-fog computing for industrial internet of things, in: WCNC 2018, 2018, pp. 1–6.

[71] X. Tao, K. Ota, M. Dong, H. Qi, K. Li, Performance guaranteed computation offloading for mobile-edge cloud computing, IEEE Wireless Commun. Letters 6 (6) (2017) 774–777.

[72] L. Mai, N. Dao, M. Park, Real-time task assignment approach leveraging reinforcement learning with evolution strategies for long-term latency minimization in fog computing, Sensors 18 (9) (2018) 2830.

[73] A. Kattepur, H. Dohare, V. Mushunuri, H. K. Rath, A. Simha, Resource constrained offloading in fog computing, in: MECC@Middleware 2016, 2016, p. 1.

[74] B. Gao, L. He, L. Liu, K. Li, S. A. Jarvis, From mobiles to clouds: Developing energy-aware offloading strategies for workflows, in: GRID 2012, 2012, pp. 139–146.

[75] D. T. Hoang, D. Niyato, P. Wang, Optimal admission control policy for mobile cloud computing hotspot with cloudlet, in: WCNC 2012, 2012, pp. 3145–3149.

[76] S. Barbarossa, S. Sardellitti, P. D. Lorenzo, Computation offloading for mobile cloud computing based on wide cross-layer optimization, in: 2013 Future Network & Mobile Summit, 2013, pp. 1–10.

[77] S. Barbarossa, S. Sardellitti, P. D. Lorenzo, Joint allocation of computation and communication resources in multiuser mobile cloud computing, in: SPAWC 2013, 2013, pp. 26–30.

[78] O. Muñoz, A. Pascual-Iserte, J. Vidal, Joint allocation of radio and computational resources in wireless application offloading, in: 2013 Future Network & Mobile Summit, 2013, pp. 1–10.

[79] T. Nishio, R. Shinkuma, T. Takahashi, N. B. Mandayam, Service-oriented heterogeneous resource sharing for optimizing service latency in mobile cloud, in: MobileCloud 2013, 2013, pp. 19–26.

[80] S. Bohez, T. Verbelen, P. Simoens, B. Dhoedt, Allocation algorithms for autonomous management of collaborative cloudlets, in: MobileCloud 2014, 2014, pp. 1–9.

[81] S. Barbarossa, P. D. Lorenzo, S. Sardellitti, Computation offloading strategies based on energy minimization under computational rate constraints, in: EuCNC 2014, 2014, pp. 1–5.

[82] Y. Zhang, D. Niyato, P. Wang, C. Tham, Dynamic offloading algorithm in intermittently connected mobile cloudlet systems, in: ICC 2014,

2014, pp. 4190–4195.

[83] O. Muñoz-Medina, A. Pascual-Iserte, J. Vidal, M. Molina, Energy-latency trade-off for multiuser wireless computation offloading, in: WCNC Workshops 2014, 2014, pp. 29–33.

[84] M. Molina, O. Muñoz, A. Pascual-Iserte, J. Vidal, Joint scheduling of communication and computation resources in multiuser wireless application offloading, in: PIMRC 2014, 2014, pp. 1093–1098.

[85] T. T. Huu, C. Tham, D. Niyato, To offload or to wait: An opportunistic offloading algorithm for parallel tasks in a mobile cloud, in: CloudCom 2014, 2014, pp. 182–189.

[86] S. Bohez, T. Verbelen, P. Simoens, B. Dhoedt, Discrete-event simulation for efficient and stable resource allocation in collaborative mobile cloudlets, Simulation Modelling Practice and Theory 50 (2015) 109–129.

[87] Y. Zhang, D. Niyato, P. Wang, Offloading in mobile cloudlet systems with intermittent connectivity, IEEE Trans. Mob. Comput. 14 (12) (2015) 2516–2529.

[88] O. Muñoz, A. Pascual-Iserte, J. Vidal, Optimization of radio and computational resources for energy efficiency in latency-constrained application offloading, IEEE Trans. Vehicular Technology 64 (10) (2015) 4738–4755.

[89] M. Al-Ayyoub, Y. Jararweh, L. A. Tawalbeh, E. Benkhelifa, A. Basalamah, Power optimization of large scale mobile cloud computing systems, in: FiCloud 2015, 2015, pp. 670–674.

[90] M. Jia, W. Liang, Z. Xu, M. Huang, Cloudlet load balancing in wireless metropolitan area networks, in: INFOCOM 2016, 2016, pp. 1–9.

[91] K. Zhang, Y. Mao, S. Leng, A. V. Vinel, Y. Zhang, Delay constrained offloading for mobile edge computing in cloud-enabled vehicular networks, in: RNDM 2016, 2016, pp. 288–294.

[92] J. Liu, Y. Mao, J. Zhang, K. B. Letaief, Delay-optimal computation task scheduling for mobile-edge computing systems, in: ISIT 2016, 2016, pp. 1451–1455.

[93] Y. Mao, J. Zhang, K. B. Letaief, Dynamic computation offloading for mobile-edge computing with energy harvesting devices, IEEE Journal on Selected Areas in Communications 34 (12) (2016) 3590–3605.

[94] H. Hong, P. Tsai, C. Hsu, Dynamic module deployment in a fog computing platform, in: APNOMS 2016, 2016, pp. 1–6.

[95] X. Chen, L. Jiao, W. Li, X. Fu, Efficient multi-user computation offloading for mobile-edge cloud computing, IEEE/ACM Trans. Netw. 24 (5) (2016) 2795–2808.

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

[97] Y. Yu, J. Zhang, K. B. Letaief, Joint subcarrier and CPU time allocation for mobile edge computing, in: GLOBECOM 2016, 2016, pp. 1–6.

[98] Y. Wang, M. Sheng, X. Wang, L. Wang, J. Li, Mobile-edge computing: Partial computation offloading using dynamic voltage scaling, IEEE Trans. Communications 64 (10) (2016) 4268–4282.

[99] C. You, K. Huang, Multiuser resource allocation for mobile-edge computation offloading, in: GLOBECOM 2016, 2016, pp. 1–6.

[100] Y. Mao, J. Zhang, S. Song, K. B. Letaief, Power-delay tradeoff in multi-user mobile-edge computing systems, in: GLOBECOM 2017, 2016, pp. 1–6.

[101] L. Wang, L. Jiao, D. Kliazovich, P. Bouvry, Reconciling task assignment and scheduling in mobile edge clouds, in: ICNP 2016, 2016, pp. 1–6.

[102] T. Chiu, W. Chung, A. Pang, Y. Yu, P. Yen, Ultra-low latency service provision in 5G fog-radio access networks, in: PIMRC 2016, 2016, pp. 1–6.

[103] C. Wang, C. Liang, F. R. Yu, Q. Chen, L. Tang, Computation offloading and resource allocation in wireless cellular networks with mobile edge computing, IEEE Trans. Wireless Communications 16 (8) (2017) 4924–4938.

[104] S. Yu, X. Wang, R. Langar, Computation offloading for mobile edge computing: A deep learning approach, in: PIMRC 2017, 2017, pp. 1–6.

[105] L. Chen, J. Xu, S. Zhou, Computation peer offloading in mobile edge computing with energy budgets, in: GLOBECOM 2017, 2017, pp. 1–6.

[106] Y. Sun, S. Zhou, J. Xu, EMM: energy-aware mobility management for mobile edge computing in ultra dense networks, IEEE Journal on Selected Areas in Communications 35 (11) (2017) 2637–2646.

[107] Z. Chang, Z. Zhou, T. Ristaniemi, Z. Niu, Energy efficient optimization for computation offloading in fog computing system, in: GLOBECOM 2017, 2017, pp. 1–6.

[108] M. Li, F. R. Yu, P. Si, H. Yao, E. Sun, Y. Zhang, Energy-efficient m2m communications with mobile edge computing in virtualized cellular networks, in: ICC 2017, 2017, pp. 1–6.

[109] Y. Cui, W. He, C. Ni, C. Guo, Z. Liu, Energy-efficient resource allocation for cache-assisted mobile edge computing, in: LCN 2017, 2017, pp. 640–648.

[110] C. You, K. Huang, H. Chae, B. Kim, Energy-efficient resource allocation for mobile-edge computation offloading, IEEE Trans. Wireless Communications 16 (3) (2017) 1397–1411.

[111] J. Guo, Z. Song, Y. Cui, Z. Liu, Y. Ji, Energy-efficient resource allocation for multi-user mobile edge computing, in: GLOBECOM 2017, 2017, pp. 1–7.

[112] P. Zhao, H. Tian, C. Qin, G. Nie, Energy-saving offloading by jointly allocating radio and computational resources for mobile edge computing, IEEE Access 5 (2017) 11255–11268.

[113] C. Wang, F. R. Yu, C. Liang, Q. Chen, L. Tang, Joint computation offloading and interference management in wireless cellular networks with mobile edge computing, IEEE Trans. Vehicular Technology 66 (8) (2017) 7432–7445.

[114] F. Wang, J. Xu, X. Wang, S. Cui, Joint offloading and computing optimization in wireless powered mobile-edge computing systems, in: ICC 2017, 2017, pp. 1–6.

[115] J. Zhang, W. Xia, Y. Zhang, Q. Zou, B. Huang, F. Yan, L. Shen, Joint offloading and resource allocation optimization for mobile edge computing, in: GLOBECOM 2017, 2017, pp. 1–6.

[116] Y. Chen, E. Sun, Y. Zhang, Joint optimization of transmission and processing delay in fog computing access networks, in: ICAIT 2017, 2017, pp. 155–158.

[117] Y. Mao, J. Zhang, K. B. Letaief, Joint task offloading scheduling and transmit power allocation for mobile-edge computing systems, in:

WCNC 2017, 2017, pp. 1–6.

[118] C. Liu, M. Bennis, H. V. Poor, Latency and reliability-aware task offloading and resource allocation for mobile edge computing, in: 2017 IEEE Globecom Workshops, 2017, pp. 1–7.

[119] S. Yi, Z. Hao, Q. Zhang, Q. Zhang, W. Shi, Q. Li, LAVEA: latency-aware video analytics on edge computing platform, in: SEC 2017, 2017, pp. 15:1–15:13.

[120] X. Zhang, Y. Mao, J. Zhang, K. B. Letaief, Multi-objective resource allocation for mobile edge computing systems, in: PIMRC 2017, 2017, pp. 1–5.

[121] X. Lyu, H. Tian, C. Sengul, P. Zhang, Multiuser joint task offloading and resource optimization in proximate clouds, IEEE Trans. Vehicular Technology 66 (4) (2017) 3435–3447.

[122] T. Q. Dinh, J. Tang, Q. D. La, T. Q. S. Quek, Offloading in mobile edge computing: Task allocation and computational frequency scaling, IEEE Trans. Communications 65 (8) (2017) 3571–3584.

[123] L. Wang, L. Jiao, J. Li, M. Mühlhäuser, Online resource allocation for arbitrary user mobility in distributed edge clouds, in: ICDCS 2017, 2017, pp. 1281–1290.

[124] S. Zhao, Y. Yang, X. Yang, W. Zhang, X. Luo, H. Qian, Online user association and computation offloading for fog-enabled D2D network, in: FWC 2017, 2017, pp. 1–6.

[125] K. Zhang, Y. Mao, S. Leng, S. Maharjan, Y. Zhang, Optimal delay constrained offloading for vehicular edge computing networks, in: ICC 2017, 2017, pp. 1–6.

[126] X. Yang, Z. Chen, K. Li, Y. Sun, H. Zheng, Optimal task scheduling in communication-constrained mobile edge computing systems for wireless virtual reality, in: APCC 2017, 2017, pp. 1–6.

[127] M. S. ElBamby, M. Bennis, W. Saad, Proactive edge computing in latency-constrained fog networks, in: EuCNC 2017, 2017, pp. 1–6.

[128] Y. Xiao, M. Krunz, QoE and power efficiency tradeoff for fog computing networks with fog node cooperation, in: INFOCOM 2017, 2017, pp. 1–9.

[129] Y. Zhou, F. R. Yu, J. Chen, Y. Kuo, Resource allocation for information-centric virtualized heterogeneous networks with in-network caching and mobile edge computing, IEEE Trans. Vehicular Technology 66 (12) (2017) 11339–11351.

[130] V. Mushunuri, A. Kattepur, H. K. Rath, A. Simha, Resource optimization in fog enabled IoT deployments, in: FMEC 2017, 2017, pp. 6–13.

[131] H. Chai, S. Leng, J. Hu, K. Yang, Resources allocation in SWIPT aided fog computing networks, in: ICAIT 2017, 2017, pp. 239–244.

[132] C. Tang, X. Wei, S. Xiao, W. Chen, W. Fang, W. Zhang, M. Hao, A mobile cloud based scheduling strategy for industrial internet of things, IEEE Access 6 (2018) 7262–7275.

[133] W. Fang, X. Yao, X. Zhao, J. Yin, N. Xiong, A stochastic control approach to maximize profit on service provisioning for mobile cloudlet platforms, IEEE Trans. Systems, Man, and Cybernetics: Systems 48 (4) (2018) 522–534.

[134] S. Bi, Y. A. Zhang, An ADMM based method for computation rate maximization in wireless powered mobile-edge computing networks, in: ICC 2018, 2018, pp. 1–7.

[135] K. Guo, M. Yang, Y. Zhang, Y. Ji, An efficient dynamic offloading approach based on optimization technique for mobile edge computing, in: MobileCloud 2018, 2018, pp. 29–36.

[136] Z. Xu, R. Gu, T. Huang, H. Xiang, X. Zhang, L. Qi, X. Xu, An IoT-Oriented offloading method with privacy preservation for cloudlet-enabled wireless metropolitan area networks, Sensors 18 (9) (2018) 3030.

[137] C. You, Y. Zeng, R. Zhang, K. Huang, Asynchronous mobile-edge computation offloading: Energy-efficient resource management, IEEE Trans. Wireless Communications 17 (11) (2018) 7590–7605.

[138] I. Lera, C. Guerrero, C. Juiz, Comparing centrality indices for network usage optimization of data placement policies in fog devices, in: FMEC 2018, 2018, pp. 115–122.

[139] L. Chen, X. Li, H. Ji, V. C. Leung, Computation offloading balance in small cell networks with mobile edge computing, Wireless Networks (2018) 1–13.

[140] N. T. Ti, L. B. Le, Computation offloading in MIMO based mobile edge computing systems under perfect and imperfect CSI estimation, in: ICC 2018, 2018, pp. 1–6.

[141] K. Guo, M. Yang, Y. Zhang, Computation offloading over a shared communication channel for mobile cloud computing, in: WCNC 2018, 2018, pp. 1–6.

[142] C. Sun, J. Zhou, J. Liuliang, J. Zhang, X. Zhang, W. Wang, Computation offloading with virtual resources management in mobile edge networks, in: VTC Spring 2018, 2018, pp. 1–5.

[143] S. Bi, Y. J. Zhang, Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading, IEEE Trans. Wireless Communications 17 (6) (2018) 4177–4190.

[144] Y. Wang, M. Sheng, X. Wang, J. Li, Cooperative dynamic voltage scaling and radio resource allocation for energy-efficient multiuser mobile edge computing, in: ICC 2018, 2018, pp. 1–6.

[145] Y. Yang, S. Zhao, W. Zhang, Y. Chen, X. Luo, J. Wang, DEBTS: delay energy balanced task scheduling in homogeneous fog networks, IEEE Internet of Things Journal 5 (3) (2018) 2094–2106.

[146] J. Li, H. Gao, T. Lv, Y. Lu, Deep reinforcement learning based computation offloading and resource allocation for MEC, in: WCNC 2018, 2018, pp. 1–6.

[147] S. Tayade, P. Rost, A. Mäder, H. D. Schotten, Delay constrained energy optimization for edge cloud offloading, in: ICC Workshops 2018, 2018, pp. 1–6.

[148] Y. Hu, A. Schmeink, Delay-constrained communication in edge computing networks, in: SPAWC 2018, 2018, pp. 1–5.

[149] X. Lyu, W. Ni, H. Tian, R. P. Liu, X. Wang, G. B. Giannakis, A. Paulraj, Distributed online optimization of fog computing for selfish devices with out-of-date information, IEEE Trans. Wireless Communications 17 (11) (2018) 7704–7717.

[150] X. Lyu, C. Ren, W. Ni, H. Tian, R. P. Liu, Distributed optimization of collaborative regions in large-scale inhomogeneous fog computing, IEEE Journal on Selected Areas in Communications 36 (3) (2018) 574–586.

[151] Y. Kim, J. Kwak, S. Chong, Dual-side optimization for cost-delay tradeoff in mobile edge computing, IEEE Trans. Vehicular Technology 67 (2) (2018) 1765–1781.

[152] F. Wang, X. Zhang, Dynamic computation offloading and resource allocation over mobile edge computing networks with energy harvesting capability, in: ICC 2018, 2018, pp. 1–6.

[153] F. Wang, X. Zhang, Dynamic interface-selection and resource allocation over heterogeneous mobile edge-computing wireless networks with energy harvesting, in: INFOCOM Workshops 2018, 2018, pp. 190–195.

[154] J. Du, L. Zhao, J. Feng, X. Chu, F. R. Yu, Economical revenue maximization in cache enhanced mobile edge computing, in: ICC 2018, 2018, pp. 1–6.

[155] A. Kiani, N. Ansari, Edge computing aware NOMA for 5G networks, IEEE Internet of Things Journal 5 (2) (2018) 1299–1306.

[156] D. T. Nguyen, L. B. Le, V. Bhargava, Edge computing resource procurement: An online optimization approach, in: WF-IoT 2018, 2018, pp. 807–812.

[157] M. Chen, Y. Hao, L. Hu, M. S. Hossain, A. Ghoneim, Edge-CoCaCo: Toward joint optimization of computation, caching, and communication on edge cloud, IEEE Wireless Commun. 25 (3) (2018) 21–27.

[158] X. Chen, W. Li, S. Lu, Z. Zhou, X. Fu, Efficient resource allocation for on-demand mobile-edge cloud computing, IEEE Trans. Vehicular Technology 67 (9) (2018) 8769–8780.

[159] S. Lagén, A. Pascual-Iserte, O. Muñoz, J. Vidal, Energy efficiency in latency-constrained application offloading from mobile clients to multiple virtual machines, IEEE Trans. Signal Processing 66 (4) (2018) 1065–1079.

[160] F. Guo, H. Zhang, H. Ji, X. Li, V. C. M. Leung, Energy efficient computation offloading for multi-access MEC enabled small cell networks, in: ICC Workshops 2018, 2018, pp. 1–6.

[161] Y. Hao, M. Chen, L. Hu, M. S. Hossain, A. Ghoneim, Energy efficient task caching and offloading for mobile edge computing, IEEE Access 6 (2018) 11365–11373.

[162] N. Nouri, A. Tadaion, Energy optimal resource allocation for mobile edge computation offloading in presence of computing access point, in: IWCIT 2018, 2018, pp. 1–6.

[163] G. Zhang, W. Zhang, Y. Cao, D. Li, L. Wang, Energy-delay tradeoff for dynamic offloading in mobile-edge computing system with energy harvesting devices, IEEE Trans. Industrial Informatics 14 (10) (2018) 4642–4655.

[164] X. Lyu, H. Tian, W. Ni, Y. Zhang, P. Zhang, R. P. Liu, Energy-efficient admission of delay-sensitive tasks for mobile edge computing, IEEE Trans. Communications 66 (6) (2018) 2603–2616.

[165] K. Cheng, Y. Teng, W. Sun, A. Liu, X. Wang, Energy-efficient joint offloading and wireless resource allocation strategy in multi-mec server systems, in: ICC 2018, 2018, pp. 1–6.

[166] M. Li, R. Yu, P. Si, Y. Zhang, Energy-efficient machine-to-machine (m2m) communications in virtualized cellular networks with mobile edge computing (mec), IEEE Transactions on Mobile Computing.

[167] H. Zhang, Z. Chen, J. Wu, Y. Deng, Y. Xiao, K. Liu, M. Li, Energy-efficient online resource management and allocation optimization in multi-user multi-task mobile-edge computing systems with hybrid energy harvesting, Sensors 18 (9) (2018) 3140.

[168] J. Zhang, X. Hu, Z. Ning, E. C. H. Ngai, L. Zhou, J. Wei, J. Cheng, B. Hu, Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks, IEEE Internet of Things Journal 5 (4) (2018) 2633–2645.

[169] C. Zhu, G. Pastor, Y. Xiao, Y. Li, A. Ylä-Jääski, Fog following me: Latency and quality balanced task allocation in vehicular fog computing, in: SECON 2018, 2018, pp. 298–306.

[170] Y. Yu, X. Bu, K. Yang, Z. Han, Green fog computing resource allocation using joint benders decomposition, dinkelbach algorithm, and modified distributed inner convex approximation, in: ICC 2018, 2018, pp. 1–6.

[171] M. Li, F. R. Yu, P. Si, Y. Zhang, Green machine-to-machine communications with mobile edge computing and wireless network virtualization, IEEE Communications Magazine 56 (5) (2018) 148–154.

[172] J. Zhang, Z. Zhou, S. Li, L. Gan, X. Zhang, L. Qi, X. Xu, W. Dou, Hybrid computation offloading for smart home automation in mobile cloud computing, Personal and Ubiquitous Computing 22 (1) (2018) 121–134.

[173] M. Zeng, Y. Li, K. Zhang, M. Waqas, D. Jin, Incentive mechanism design for computation offloading in heterogeneous fog computing: A contract-based approach, in: ICC 2018, 2018, pp. 1–6.

[174] S. Seng, X. Li, H. Ji, H. Zhang, Joint access selection and heterogeneous resources allocation in udns with MEC based on non-orthogonal multiple access, in: ICC Workshops 2018, 2018, pp. 1–6.

[175] M. Ali, N. Riaz, M. I. Ashraf, S. B. Qaisar, M. Naeem, Joint cloudlet selection and latency minimization in fog networks, IEEE Trans. Industrial Informatics 14 (9) (2018) 4055–4063.

[176] X. Cao, F. Wang, J. Xu, R. Zhang, S. Cui, Joint computation and communication cooperation for mobile edge computing, in: WiOpt 2018, 2018, pp. 1–6.

[177] J. Zhang, W. Xia, F. Yan, L. Shen, Joint computation offloading and resource allocation optimization in heterogeneous networks with mobile edge computing, IEEE Access 6 (2018) 19324–19337.

[178] S. Mu, Z. Zhong, D. Zhao, M. Ni, Joint job partitioning and collaborative computation offloading for internet of things, IEEE Internet of Things Journal.

[179] F. Wang, J. Xu, X. Wang, S. Cui, Joint offloading and computing optimization in wireless powered mobile-edge computing systems, IEEE Trans. Wireless Communications 17 (3) (2018) 1784–1797.

[180] M. Guan, B. Bai, L. Wang, S. Jin, Z. Han, Joint optimization for computation offloading and resource allocation in internet of things, in: VTC Fall 2017, 2017, pp. 1–5.

[181] L. Cui, C. Xu, S. Yang, J. Z. Huang, J. Li, X. Wang, Z. Ming, N. Lu, Joint optimization of energy consumption and latency in mobile edge computing for internet of things, IEEE Internet of Things Journal.

[182] Y. Yang, Y. Ma, W. Xiang, X. Gu, H. Zhao, Joint optimization of energy consumption and packet scheduling for mobile edge computing in cyber-physical networks, IEEE Access 6 (2018) 15576–15586.

[183] Y. Gu, Z. Chang, M. Pan, L. Song, Z. Han, Joint radio and computational resource allocation in IoT fog computing, IEEE Trans. Vehicular Technology 67 (8) (2018) 7475–7484.

[184] H. Xing, L. Liu, J. Xu, A. Nallanathan, Joint task assignment and wireless resource allocation for cooperative mobile-edge computing, in: ICC 2018, 2018, pp. 1–6.

[185] J. Ren, G. Yu, Y. Cai, Y. He, Latency optimization for resource allocation in mobile-edge computation offloading, IEEE Trans. Wireless Communications 17 (8) (2018) 5506–5519.
[186] T.-C. Chiu, A.-C. Pang, W.-H. Chung, J. Zhang, Latency-driven fog cooperation approach in fog radio access networks, IEEE Transactions on Services Computing.
[187] G. Chi, Y. Wang, X. Liu, Y. Qiu, Latency-optimal task offloading for mobile-edge computing system in 5G heterogeneous networks, in: VTC Spring 2018, 2018, pp. 1–5.
[188] Y. Yang, K. Wang, G. Zhang, X. Chen, X. Luo, M. Zhou, MEETS: maximal energy efficient task scheduling in homogeneous fog networks, IEEE Internet of Things Journal 5 (5) (2018) 4076–4087.
[189] X. Yang, Z. Liu, Y. Yang, Minimization of weighted bandwidth and computation resources of fog servers under per-task delay constraint, in: ICC 2018, 2018, pp. 1–6.
[190] L. Yang, H. Zhang, M. Li, J. Guo, H. Ji, Mobile edge computing empowered energy efficient task offloading in 5G, IEEE Trans. Vehicular Technology 67 (7) (2018) 6398–6409.
[191] X. Cao, J. Xu, R. Zhang, Mobile edge computing for cellular-connected UAV: computation offloading and trajectory optimization, in: SPAWC 2018, 2018, pp. 1–5.
[192] J. Li, A. Wu, S. Chu, T. Liu, F. Shu, Mobile edge computing for task offloading in small-cell networks via belief propagation, in: ICC 2018, 2018, pp. 1–6.
[193] S. Jeong, O. Simeone, J. Kang, Mobile edge computing via a UAV-mounted cloudlet: Optimization of bit allocation and path planning, IEEE Trans. Vehicular Technology 67 (3) (2018) 2049–2063.
[194] L. Jiao, L. Pu, L. Wang, X. Lin, J. Li, Multiple granularity online control of cloudlet networks for edge computing, in: SECON 2018, 2018, pp. 406–414.
[195] W. Chen, D. Wang, K. Li, Multi-user multi-task computation offloading in green mobile edge cloud computing, IEEE Transactions on Services Computing.
[196] X. Wang, Z. Ning, L. Wang, Offloading in internet of vehicles: A fog-enabled real-time traffic management system, IEEE Trans. Industrial Informatics 14 (10) (2018) 4568–4578.
[197] J. Liu, Q. Zhang, Offloading schemes in mobile edge computing for ultra-reliable low latency communications, IEEE Access 6 (2018) 12825–12837.
[198] H. Wu, L. Chen, C. Shen, W. Wen, J. Xu, Online geographical load balancing for energy-harvesting mobile edge computing, in: ICC 2018, 2018, pp. 1–6.
[199] S. Sardellitti, M. Merluzzi, S. Barbarossa, Optimal association of mobile users to multi-access edge computing resources, in: ICC Workshops 2018, 2018, pp. 1–6.
[200] Z. Wei, H. Jiang, Optimal offloading in fog computing systems with non-orthogonal multiple access, IEEE Access 6 (2018) 49767–49778.
[201] N. Mohan, J. Kangasharju, Placing it right!: optimizing energy, processing, and transport in edge-fog clouds, Annales des Télécommunications 73 (7-8) (2018) 463–474.
[202] X. Hu, K. Wong, K. Yang, Power minimization for cooperative wireless powered mobile edge computing systems, in: ICC 2018, 2018, pp. 1–6.
[203] X. Chen, L. Wang, C. Wang, R. Jin, Predictive offloading in mobile-fog-cloud enabled cyber-manufacturing systems, in: IEEE Industrial Cyber-Physical Systems, 2018, pp. 167–172.
[204] L. Liu, Q. Fan, Resource allocation optimization based on mixed integer linear programming in the multi-cloudlet environment, IEEE Access 6 (2018) 24533–24542.
[205] C. You, Y. Zeng, R. Zhang, K. Huang, Resource management for asynchronous mobile-edge computation offloading, in: ICC Workshops 2018, 2018, pp. 1–6.
[206] Z.-z. Liu, S.-n. Li, Sensor-cloud data acquisition based on fog computation and adaptive block compressed sensing, International Journal of Distributed Sensor Networks 14 (9) (2018) 1550147718802259.
[207] M. Chen, Y. Hao, Task offloading for mobile edge computing in software defined ultra-dense network, IEEE Journal on Selected Areas in Communications 36 (3) (2018) 587–597.
[208] Q. Fan, N. Ansari, Towards workload balancing in fog computing empowered IoT, IEEE Transactions on Network Science and Engineering.
[209] F. Zhou, Y. Wu, H. Sun, Z. Chu, UAV-enabled mobile edge computing: Offloading optimization and trajectory design, in: ICC 2018, 2018, pp. 1–6.
[210] Z. Wang, Z. Zhong, D. Zhao, M. Ni, Vehicle-based cloudlet relaying for mobile computation offloading, IEEE Trans. Vehicular Technology 67 (11) (2018) 11181–11191.
[211] X. Hu, K. Wong, K. Yang, Wireless powered cooperation-assisted mobile edge computing, IEEE Trans. Wireless Communications 17 (4) (2018) 2375–2388.
[212] Y. Zhang, X. Chen, Y. Chen, Z. Li, J. Huang, Cost efficient scheduling for delay-sensitive tasks in edge computing system, in: SCC 2018, 2018, pp. 73–80.
[213] D. Rahbari, S. Kabirzadeh, M. Nickray, A security aware scheduling in fog computing by hyper heuristic algorithm, in: ICSPIS 2017, 2017, pp. 87–92.
[214] S. Li, J. Huang, Energy efficient resource management and task scheduling for IoT services in edge computing paradigm, in: ISPA/IUCC 2017, 2017, pp. 846–851.
[215] H. Tan, Z. Han, X. Li, F. C. M. Lau, Online job dispatching and scheduling in edge-clouds, in: INFOCOM 2017, 2017, pp. 1–9.
[216] X. Lyu, W. Ni, H. Tian, R. P. Liu, X. Wang, G. B. Giannakis, A. Paulraj, Optimal schedule of mobile edge computing for internet of things using partial information, IEEE Journal on Selected Areas in Communications 35 (11) (2017) 2606–2615.
[217] X. Wang, K. Wang, S. Wu, S. Di, H. Jin, K. Yang, S. Ou, Dynamic resource scheduling in mobile edge cloud with cloud radio access network, IEEE Trans. Parallel Distrib. Syst. 29 (11) (2018) 2429–2445.
[218] T. Wang, X. Wei, C. Tang, J. Fan, Efficient multi-tasks scheduling algorithm in mobile cloud computing with time constraints, Peer-to-Peer Networking and Applications 11 (4) (2018) 793–807.

[219] J. Wan, B. Chen, S. Wang, M. Xia, D. Li, C. Liu, Fog computing for energy-aware load balancing and scheduling in smart factory, IEEE Trans. Industrial Informatics 14 (10) (2018) 4548–4556.

[220] Z. Liu, J. Zhang, Y. Li, L. Bai, Y. Ji, Joint jobs scheduling and lightpath provisioning in fog computing micro datacenter networks, Journal of Optical Communications and Networking 10 (7) (2018) B152–B163.

[221] Y. Yang, K. Wang, G. Zhang, X. Chen, X. Luo, M. Zhou, Maximal energy efficient task scheduling for homogeneous fog networks, in: INFOCOM Workshops 2018, 2018, pp. 274–279.

[222] Y. Sun, F. Lin, H. Xu, Multi-objective optimization of resource scheduling in fog computing using an improved NSGA-II, Wireless Personal Communications 102 (2) (2018) 1369–1385.

[223] S. K. Mishra, D. Puthal, J. J. P. C. Rodrigues, B. Sahoo, E. Dutkiewicz, Sustainable service allocation using a metaheuristic technique in a fog server for industrial applications, IEEE Trans. Industrial Informatics 14 (10) (2018) 4497–4506.

[224] K. Lin, S. Pankaj, D. Wang, Task offloading and resource allocation for edge-of-things computing on smart healthcare systems, Computers & Electrical Engineering 72 (2018) 348–360.

[225] J. Luo, X. Deng, H. Zhang, H. Qi, Ultra-low latency service provision in edge computing, in: ICC 2018, 2018, pp. 1–6.

[226] J. Oueis, E. C. Strinati, S. Barbarossa, Small cell clustering for efficient distributed cloud computing, in: PIMRC 2014, 2014, pp. 1474–1479.

[227] J. Oueis, E. C. Strinati, S. Sardellitti, S. Barbarossa, Small cell clustering for efficient distributed fog computing: A multi-user case, in: VTC Fall 2015, 2015, pp. 1–5.

[228] J. Oueis, E. C. Strinati, S. Barbarossa, The fog balancing: Load distribution for small cell cloud computing, in: VTC Spring 2015, 2015, pp. 1–6.

[229] G. Lee, W. Saad, M. Bennis, An online secretary framework for fog network formation with minimal latency, in: ICC 2017, 2017, pp. 1–6.

[230] E. Balevi, R. D. Gitlin, A clustering algorithm that maximizes throughput in 5G heterogeneous F-RAN networks, in: ICC 2018, 2018, pp. 1–6.

[231] D. Kimovski, H. Ijaz, N. Saurabh, R. Prodan, Adaptive nature-inspired fog architecture, in: ICFEC 2018, 2018, pp. 1–8.

[232] Y. Liu, F. R. Yu, X. Li, H. Ji, H. Zhang, V. C. M. Leung, Joint access and resource management for delay-sensitive transcoding in ultra-dense networks with mobile edge computing, in: ICC 2018, 2018, pp. 1–6.

[233] S. Wang, R. Urgaonkar, M. Zafer, T. He, K. S. Chan, K. K. Leung, Dynamic service migration in mobile edge-clouds, in: Networking 2015, 2015, pp. 1–9.

[234] H. Yao, C. Bai, D. Zeng, Q. Liang, Y. Fan, Migrate or not? exploring virtual machine migration in roadside cloudlet-based vehicular cloud, Concurrency and Computation: Practice and Experience 27 (18) (2015) 5780–5792.

[235] D. Zhao, T. Yang, Y. Jin, Y. Xu, A service migration strategy based on multiple attribute decision in mobile edge computing, in: ICCT 2017, 2017, pp. 986–990.

[236] Y. Chen, J. P. Walters, S. P. Crago, Load balancing for minimizing deadline misses and total runtime for connected car systems in fog computing, in: ISPA/IUCC 2017, 2017, pp. 683–690.

[237] T. G. Rodrigues, K. Suto, H. Nishiyama, N. Kato, K. Temma, Cloudlets activation scheme for scalable mobile edge computing with transmission power control and virtual machine migration, IEEE Trans. Computers 67 (9) (2018) 1287–1300.

[238] T. Ouyang, Z. Zhou, X. Chen, Follow me at the edge: Mobility-aware dynamic service placement for mobile edge computing, IEEE Journal on Selected Areas in Communications 36 (10) (2018) 2333–2345.

[239] L. Wang, L. Jiao, J. Li, J. Gedeon, M. Mühlhäuser, Moera: Mobility-agnostic online resource allocation for edge computing, IEEE Transactions on Mobile Computing.

[240] Z. Xu, W. Liang, W. Xu, M. Jia, S. Guo, Capacitated cloudlet placements in wireless metropolitan area networks, in: LCN 2015, 2015, pp. 570–578.

[241] A. Ceselli, M. Premoli, S. Secci, Cloudlet network design optimization, in: Networking 2015, 2015, pp. 1–9.

[242] Z. Xu, W. Liang, W. Xu, M. Jia, S. Guo, Efficient algorithms for capacitated cloudlet placements, IEEE Trans. Parallel Distrib. Syst. 27 (10) (2016) 2866–2880.

[243] A. Ceselli, M. Premoli, S. Secci, Heuristics for static cloudlet location, Electronic Notes in Discrete Mathematics 55 (2016) 21–24.

[244] C. Wang, S. Zhang, H. Zhang, Z. Qian, S. Lu, Edge cloud capacity allocation for low delay computing on mobile devices, in: ISPA/IUCC 2017, 2017, pp. 290–297.

[245] P. Maiti, J. Shukla, B. Sahoo, A. K. Turuk, Efficient data collection for IoT services in edge computing environment, in: ICIT 2017, 2017, pp. 101–106.

[246] A. Ceselli, M. Premoli, S. Secci, Mobile edge cloud network design optimization, IEEE/ACM Trans. Netw. 25 (3) (2017) 1818–1831.

[247] I. Gravalos, P. Makris, K. Christodoulopoulos, E. A. Varvarigos, Efficient network planning for internet of things with QoS constraints, IEEE Internet of Things Journal 5 (5) (2018) 3823–3836.

[248] G. Premsankar, B. Ghaddar, M. D. Francesco, R. Verago, Efficient placement of edge computing devices for vehicular applications in smart cities, in: 2018 IEEE/IFIP Network Operations and Management Symposium, 2018, pp. 1–9.

[249] A. Santoyo-González, C. Cervelló-Pastor, Latency-aware cost optimization of the service infrastructure placement in 5G networks, J. Network and Computer Applications 114 (2018) 29–37.

[250] M. Bouet, V. Conan, Mobile edge computing resources optimization: A geo-clustering approach, IEEE Trans. Network and Service Management 15 (2) (2018) 787–796.

[251] L. Zhao, W. Sun, Y. Shi, J. Liu, Optimal placement of cloudlets for access delay minimization in sdn-based internet of things networks, IEEE Internet of Things Journal 5 (2) (2018) 1334–1344.

[252] E. Balevi, R. D. Gitlin, Optimizing the number of fog nodes for cloud-fog-thing networks, IEEE Access 6 (2018) 11173–11183.

[253] L. Chen, J. Wu, G. Zhou, L. Ma, QUICK: QoS-guaranteed efficient cloudlet placement in wireless metropolitan area networks, The Journal of Supercomputing 74 (8) (2018) 4037–4059.

[254] S. Guo, D. Wu, H. Zhang, D. Yuan, Resource modeling and scheduling for mobile edge computing: A service providers perspective, IEEE Access 6 (2018) 35611–35623.

[255] A. M. Haubenwaller, K. Vandikas, Computations on the edge in the internet of things, Procedia Computer Science 52 (2015) 29–34.

[256] A. M. Haubenwaller, K. Vandikas, Computations on the edge in the internet of things, in: ANT-2015, SEIT-2015, 2015, pp. 29–34.
[257] L. Gu, D. Zeng, S. Guo, A. Barnawi, Y. Xiang, Cost efficient resource management in fog computing supported medical cyber-physical system, IEEE Trans. Emerging Topics Comput. 5 (1) (2017) 108–119.
[258] D. Zeng, L. Gu, S. Guo, Z. Cheng, S. Yu, Joint optimization of task scheduling and image placement in fog computing supported software-defined embedded system, IEEE Trans. Computers 65 (12) (2016) 3702–3712.
[259] V. Cardellini, V. Grassi, F. L. Presti, M. Nardelli, Optimal operator placement for distributed stream processing applications, in: DEBS 2016, 2016, pp. 69–80.
[260] A. A. Aakizadeh, A. A. Aashi, Distribution of virtual devices on the fog for delay and araffic aeduction, in: KBEI 2017, 2017, pp. 0492–0496.
[261] X. Wang, S. Leng, X. Liu, Q. Zhao, K. Wang, K. Yang, Fog computing aided multi-view video in mobile social networks, in: ICAIT 2017, 2017, pp. 361–366.
[262] H. R. Arkian, A. Diyanat, A. Pourkhalili, MIST: fog-based data analytics scheme with cost-efficient resource provisioning for IoT crowd-sensing applications, J. Network and Computer Applications 82 (2017) 152–165.
[263] V. Karagiannis, A. Papageorgiou, Network-integrated edge computing orchestrator for application placement, in: CNSM 2017, 2017, pp. 1–5.
[264] S. Wang, M. Zafer, K. K. Leung, Online placement of multi-component applications in edge computing environments, IEEE Access 5 (2017) 2514–2533.
[265] J. Sheu, Y. Pu, R. B. Jagadeesha, Y. Chang, An efficient module deployment algorithm in edge computing, in: WCNC 2018 Workshops, 2018, pp. 208–213.
[266] Y. Xia, X. Etchevers, L. Letondeur, T. Coupaye, F. Desprez, Combining hardware nodes and software components ordering-based heuristics for optimizing the placement of distributed IoT applications in the fog, in: SAC 2018, 2018, pp. 751–760.
[267] P. Smet, B. Dhoedt, P. Simoens, Docker layer placement for on-demand provisioning of services on edge clouds, IEEE Trans. Network and Service Management 15 (3) (2018) 1161–1174.
[268] B. Jia, H. Hu, Y. Zeng, T. Xu, Y. Yang, Double-matching resource allocation strategy in fog computing networks based on cost efficiency, Journal of Communications and Networks 20 (3) (2018) 237–246.
[269] W. Kim, S. Chung, User incentive model and its optimization scheme in user-participatory fog computing environment, Computer Networks 145 (2018) 76–88.
[270] W. Kim, S. Chung, User-participatory fog computing architecture and its management schemes for improving feasibility, IEEE Access 6 (2018) 20262–20278.
[271] N. Akhtar, I. Matta, A. Raza, L. Goratti, T. Braun, F. Esposito, Virtual function placement and traffic steering over 5G multi-technology networks, in: NetSoft 2018, 2018, pp. 114–122.
[272] O. Skarlat, M. Nardelli, S. Schulte, M. Borkowski, P. Leitner, Optimized IoT service placement in the fog, Service Oriented Computing and Applications 11 (4) (2017) 427–443.
[273] O. Skarlat, S. Schulte, M. Borkowski, P. Leitner, Resource provisioning for IoT services in the fog, in: SOCA 2016, 2016, pp. 32–39.
[274] T. Hou, G. Feng, S. Qin, W. Jiang, Proactive content caching by exploiting transfer learning for mobile edge computing, in: GLOBECOM 2017, 2017, pp. 1–6.
[275] T. Hou, G. Feng, S. Qin, W. Jiang, Proactive content caching by exploiting transfer learning for mobile edge computing, Int. J. Communication Systems 31 (11).
[276] L. Zhao, J. Liu, Optimal placement of virtual machines for supporting multiple applications in mobile edge networks, IEEE Trans. Vehicular Technology 67 (7) (2018) 6533–6545.
[277] S. Wang, R. Urgaonkar, T. He, K. Chan, M. Zafer, K. K. Leung, Dynamic service placement for mobile micro-clouds with predicted future costs, IEEE Trans. Parallel Distrib. Syst. 28 (4) (2017) 1002–1016.
[278] C. T. Do, N. H. Tran, C. Pham, M. G. R. Alam, J. H. Son, C. S. Hong, A proximal algorithm for joint resource allocation and minimizing carbon footprint in geo-distributed fog computing, in: ICOIN 2015, 2015, pp. 324–329.
[279] F. B. Jemaa, G. Pujolle, M. Pariente, QoS-Aware VNF placement optimization in edge-central carrier cloud architecture, in: GLOBECOM 2016, 2016, pp. 1–7.
[280] L. Vigneri, T. Spyropoulos, C. Barakat, Storage on wheels: Offloading popular contents through a vehicular cloud, in: WoWMoM 2016, 2016, pp. 1–9.
[281] W. Tärneberg, A. V. Papadopoulos, A. Mehta, J. Tordsson, M. Kihl, Distributed approach to the holistic resource management of a mobile cloud network, in: ICFEC 2017, 2017, pp. 51–60.
[282] T. Bahreini, D. Grosu, Efficient placement of multi-component applications in edge computing systems, in: SEC 2017, 2017, pp. 5:1–5:11.
[283] O. Ascigil, T. K. Phan, A. G. Tasiopoulos, V. Sourlas, I. Psaras, G. Pavlou, On uncoordinated service placement in edge-clouds, in: CloudCom 2017, 2017, pp. 41–48.
[284] O. Skarlat, M. Nardelli, S. Schulte, S. Dustdar, Towards QoS-Aware fog service placement, in: ICFEC 2017, 2017, pp. 89–96.
[285] J. Chakareski, VR/AR immersive communication: Caching, edge computing, and transmission trade-offs, in: VR/AR Network@SIGCOMM 2017, 2017, pp. 36–41.
[286] C. Guerrero, I. Lera, C. Juiz, A lightweight decentralized service placement policy for performance optimization in fog computing, Journal of Ambient Intelligence and Humanized Computing (2018) 1–18.
[287] S. Venticinque, A. Amato, A methodology for deployment of IoT application in fog, Journal of Ambient Intelligence and Humanized Computing (2018) 1–22.
[288] T. Tran, D. Pompili, Adaptive bitrate video caching and processing in mobile-edge computing networks, IEEE Transactions on Mobile Computing.
[289] R. Ghosh, S. P. R. Komma, Y. Simmhan, Adaptive energy-aware scheduling of dynamic event analytics across edge and cloud resources, in: CCGRID 2018, 2018, pp. 72–82.
[290] L. Chen, P. Zhou, L. Gao, J. Xu, Adaptive fog configuration for the industrial internet of things, IEEE Trans. Industrial Informatics 14 (10) (2018) 4656–4664.

[291] S. Deng, Z. Xiang, J. Yin, J. Taheri, A. Y. Zomaya, Composition-driven IoT service provisioning in distributed edges, IEEE Access 6 (2018) 54258–54269.

[292] Z. Li, R. Xie, Q. Jia, T. Huang, Energy-efficient joint caching and transcoding for HTTP adaptive streaming in 5G networks with mobile edge computing, in: ICC Workshops 2018, 2018, pp. 1–6.

[293] S. Zhao, Y. Yang, Z. Shao, X. Yang, H. Qian, C. Wang, FEMOS: fog-enabled multitier operations scheduling in dynamic wireless networks, IEEE Internet of Things Journal 5 (2) (2018) 1169–1183.

[294] Y. He, N. Zhao, H. Yin, Integrated networking, caching, and computing for connected vehicles: A deep reinforcement learning approach, IEEE Trans. Vehicular Technology 67 (1) (2018) 44–55.

[295] G. Zhong, J. Yan, L. Kuang, Qoe-driven social aware caching placement for terrestrial-satellite networks, China Communications 15 (10) (2018) 60–72.

[296] P. Sermpezis, T. Giannakas, T. Spyropoulos, L. Vigneri, Soft cache hits: Improving performance through recommendation and delivery of related content, IEEE Journal on Selected Areas in Communications 36 (6) (2018) 1300–1313.

[297] Z. Tan, F. R. Yu, X. Li, H. Ji, V. C. M. Leung, Virtual resource allocation for heterogeneous services in full duplex-enabled scns with mobile edge computing and caching, IEEE Trans. Vehicular Technology 67 (2) (2018) 1794–1808.