Future Generation Computer Systems 88 (2018) 644–662

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Augmenting computing capabilities at the edge by jointly exploiting mobile devices: A survey

Matías Hirsch *, Cristian Mateos, Alejandro Zunino

ISISTAN-UNICEN-CONICET Research Institute. Campus Universitario, Tandil (B7001BBO), Argentina

HIGHLIGHTS

- Novel taxonomies inspired by Smart Mobile Devices (SMD) singularities.
- Taxonomies stem from singularity awareness level and evaluation support dimensions.
- A survey of SMD-centric resource allocation at the edge for intensive tasks.
- An outline of research challenges in the area and open future research opportunities.

ARTICLE INFO

Article history:
Received 1 December 2017
Received in revised form 27 April 2018
Accepted 2 June 2018
Available online 7 June 2018

Keywords:
Mobile cloud computing
Edge computing
Smart mobile devices
Resource-intensive applications
Resource allocation

ABSTRACT

The ever-growing adoption of smart mobile devices is a worldwide phenomenon that positions smartphones and tablets as primary devices, i.e., that people mostly use. In addition to this, the computing capabilities of such devices, often under-utilized by their owners, are in continuous improvement. Today, smart mobile devices have multi-core CPUs, several gigabytes of RAM, and the ability to communicate through several wireless networking technologies. These facts caught the attention of researchers who propose to leverage smart mobile devices aggregated computing capabilities for running resource intensive software at the edge of the network. Such idea is conditioned by key features, named singularities in the context of this work, that makes smart mobile devices resource exploitation a difficult problem. These are the ability of devices to change location (user mobility), the shared condition -i.e., non-dedicated nature- of resources provided (lack of ownership) and the limited operation time given by the finite energy source (exhaustible resources). In this paper, we provide an in-depth analysis of proposals materializing this idea. We show that (a) existing approaches differ in the singularities combinations they target and the way they address each singularity through novel taxonomies, and (b) this fact makes them suitable for distinct goals and resource exploitation opportunities also schematized in this paper. The latter are represented by real life situations where resources provided by groups of smart mobile devices can be exploited, which in turn, are characterized by a social context and a networking support used to link and coordinate devices. The behavior of people in a given social context configure a special availability level of resources, while the networking support imposes restrictions on how data/computational tasks are distributed and results are collected. We conclude our analysis by discussing strong/weak points of the approaches and by identifying prospective future lines in the area.

1. Introduction

With the evolution of the semiconductor industry, cellphones, which were used mainly for communication and Internet access, were progressively equipped with more and more powerful components. CPUs with multiple cores, GPU, several gigabytes of main memory and storage space, a variety sensors, a high resolution screen and cameras, and multiple wireless data transferring interfaces are assembled in what is commonly known as *smartphone*. Tablets are smartphone-like devices equipped with higher screens and usually less communication technologies. Because of the rich capabilities derived from the combination of all aforementioned components, smartphones and tablets are commonly called *Smart Mobile Devices* (SMDs). This acronym as well as "device" and "node" will be used interchangeably in the rest of the document to refer to such type of devices.

Every new SMD generation surpasses the previous one, or even previous generations of personal computers, in respect to

* Corresponding author.
E-mail addresses: matias.hirsch@isistan.unicen.edu.ar (M. Hirsch), cristian.mateos@isistan.unicen.edu.ar (C. Mateos), alejandro.zunino@isistan.unicen.edu.ar (A. Zunino).
URLs: http://www.exa.unicen.edu.ar/∼cmateos (C. Mateos), http://www.exa.unicen.edu.ar/∼azunino (A. Zunino).

https://doi.org/10.1016/j.future.2018.06.005

type of applications they can execute. Such capability has been welcomed and exploited in professional areas including, but not limited to, healthcare [1–3], engineering [4], civil construction [5] and agriculture [6], which rely on mobile applications as a cheap approach to aid in-situ professional tasks. The existence of the BOINC for Android application is another example that evidences the power of SMDs computing capabilities, in this case related to scientific research. Applications like these make heavy use of SMDs computing resources – CPU, GPU, memory – to execute pattern recognition, image and signal processing techniques, complex numerical methods and machine learning algorithms.

Despite SMDs are powerful enough to execute such applications, the limited energy might prevent applications from terminating successfully, or executing with the user desired QoS. Notably, offloading [7,8] is a technique which aims at mitigating such limitations by delegating the execution of certain application software components to remote resource rich infrastructures like Clouds, a paradigm known with the name of Mobile Cloud Computing (MCC) [9,10]. MCC research endeavors have been focused on the study of software architectural features of mobile applications to ease the delegation of code execution [11], study decision making components for evaluating energy efficiency of local (mobile) versus remote (Cloud, Cloudlet) component execution. It is also challenging to achieve a seamless usage of remote resources [12] considering, for instance, SMDs connection quality while moving through different Cloudlets [13].

Aware of SMDs energy constraints, and at the same time conscious of the worldwide popularity and the continuous improvement of SMD computing capabilities, alternative research lines consider SMDs as first-class computing providers. More specifically, instead of using resource-rich infrastructures to mitigate SMDs' individual limitations, such alternative efforts are aligned in understanding how heterogeneous resources of a group of SMDs can be collectively exploited to execute resource intensive software. The initiatives which aim at studying and evaluating technical aspects concerning the integration of SMDs into distributed and high performance computing platforms [14–18], are examples of such efforts. Other efforts in the same line envision local collaborations among nearby SMDs to accomplish delay-sensitive tasks and/or as a way of augmenting SMDs individual capabilities in cases where accessing remote computing resources is not a convenient or viable option [19–21]. Moreover, since SMDs might be close to "things" that produce/sense data, frameworks for executing in-the-field applications which aim at scavenging SMDs' computing/communicating resources are also studied in the context of IoT applications [22–24]. Taking advantage of SMDs' resources locally would help to alleviate the usage of power-hungry servers and reduce the injection of enormous amounts of data traffic to the Internet.

For coordinating the usage of resources in a group of SMDs, it is necessary to adopt an underlying networking support, i.e., typically an ad-hoc [25,26] or an infrastructure-based [27] network. Besides, to make an effective use of the available resources, it is necessary to provide a resource allocation (RA) mechanism designed to deal with resource heterogeneity and dynamic availability not quite present in traditional distributed computing environments [28–31], which poses challenges. These challenges, which stems from the ability of SMDs to change location, their non-dedicated nature of resources and the limited operation time imposed by the finite energy supply, hereafter *SMDs' singularities*, are developed in Section 3.2.

In this survey, we describe, analyze and classify RA mechanisms that consider SMDs as first-class computing providers. Our vision to classify works comprises an original multi-singularity driven study that we propose as starting point framework for organizing current and future RA efforts in the area. As discussed in Section 3.2, SMDs singularities represent underpinning views for the design of RA mechanisms aiming at exploiting SMDs resources. Providing that these concepts can be more or less present in different social contexts, where groups of SMDs are expected to be found, a first-cut categorization and the notion of *resource exploitation opportunity* is introduced.

In the literature, several RA mechanisms targeting the exploitation of computing resources of SMDs have been proposed. These mechanisms are somewhat focused on the challenges posed by one or two but not all of the singularities. Based on this, we propose a first-cut classification of works and relate the influence/non-influence of a singularity to the notion of *resource exploitation context*. We conceive the latter as an abstract concept where the dynamics of a social context, the characteristics of a networking support and the requirements of a compute-intensive software confluence. We take advantage of this first classification to describe aspects of RA works that are not discussed in any other section of the survey.

A second-cut analysis of works shows how singularities – i.e., the challenges associated – are addressed. We identified RA mechanisms which rely on specific information and/or propose concrete actions from the RA logic to deal with a singularity, so we tagged them as *aware of* the singularity. Other mechanisms, by contrast, provide an evaluation of inherited/classical RA mechanisms under experimental scenarios associated to a singularity, e.g., nodes movement. We tagged these efforts as *evaluated for* a singularity. Both types of works, i.e., aware of and evaluated for a singularity, adopt an implementation scheme for reproducing the experimental conditions associated to a singularity. We identified in-vivo and simulated methodologies related to this concern.

Our three-singularity driven study delineated to classify works aims at being an starting point framework for organizing current and future RA efforts in the area. The rest of this document is structured as follows. The next section positions the contributions of this survey in light of other surveys already published in related research areas. Section 3 describes preliminary notions related to the RA problem in groups of SMDs – in this survey mobile Grids –, including a conceptual overview of the most typical forms of connecting SMDs and the singularities that characterize resource provision with SMDs. Section 4 describes the state-of-the-art RA mechanisms classified according to the combination of singularities they address. Section 5 steps into details about how singularities have been addressed by existing RA works through novel taxonomies and concludes with a discussion that relates the efforts aligned with each singularity, the type of networking support used and the evaluation methodology employed. Section 6 lists future directions that researchers in the area would need to consider. Section 7 exposes the conclusions of the survey.

2. Related works and contribution

Mobile-oriented Cloud architectures and technologies undoubtedly play a very important role in practice due to the massive adoption of mobile devices by users worldwide. From the industry side, the synergy between mobile and Cloud technologies has produced brand new Cloud provisioning models for supporting mobile application development and deployment, such as mobile backend as a service (MBaaS). MBaaS essentially identifies Cloud services which are commonly needed by Web and mobile applications such as data storage, identity and access management, synchronization, push notifications, access to third-party services – e.g. social networks – via API gateways, and so on. From an academic perspective, the Mobile Cloud Computing (MCC) paradigm [10] has been proposed as a way to augment mobile devices – and hence deal with their inherent limitations – with remote resources located in the Cloud. To this end, MCC combines advances from the areas of mobile computing, cloud computing and wireless/fixed networks so that rich applications can be seamlessly and efficiently "executed" in mobile devices via the actual execution/processing of computations/data on remote Cloud resources. These rich applications might include for instance speech recognition, augmented reality, and many more.

Naturally, MCC is not free from research challenges. Perhaps the most evident one and the closest to the purpose of this survey is how and when to move computations/data from mobile devices to remote resources while maximizing application execution performance and energy saving. This problem is referred as offloading [10,32] or cyber-foraging [33]. At its core, the problem involves approaches to properly analyze under which situations using offloading is reasonable or not. The canonical model for evaluating the trade-off between computing locally or remotely takes the execution time of the computation to be performed (TE), the amount of executable code and input data to transmit (DT), the number of bytes used to receive the result (DR), the transmission bandwidth (BW) and the bandwidth of the reception network (BR). Then, $T_{offloading} = \frac{DR}{BW} + \frac{DR}{BR} + TE$ shows that offloading is convenient when the estimated execution time on the mobile device is higher than $T_{offloading}$ [33].

Literature evidences many research efforts proposing elaborated offloading techniques to cope with several aspects of the problem. For example, taking good offloading decisions depends on various parameters, namely application characteristics, network conditions, hardware features and target operating systems. Then, the survey in [34] discusses existing techniques to deal with variations in these parameter values by designing effective offloading solutions that can adapt gracefully to changes. The survey in [35] focuses on analyzing offloading techniques from the perspective of speedup (computing speed gains) rather than energy saving in mobile devices. Moreover, another recent survey reviews Cloud-based frameworks for *seamless* execution of mobile applications on the Cloud [12], i.e., techniques allowing users to augment mobile device capabilities with as little user involvement, interaction and distraction as possible. Challenges involved are transparent Cloud discovery, service selection and application components deployment.

Traditional MCC was somewhere in the middle of the path claimed to be insufficient for many latency-sensitive applications and very large number of (mobile) client devices, such as critical IoT services. Motivated by this fact, researchers from Cisco introduced in 2012 the Fog Computing paradigm [36]. Fog Computing focuses on providing highly-scalable infrastructures for latency and location-aware applications, where geographical distribution, mobility and software/hardware heterogeneity prevail. While Fog Computing can be viewed as a special case of MCC, it represents at the same time an evolution of the latter since a distinctive aspect is the ability of augmenting mobile (e.g., laptops, smartphones, tablets, wearables) and wireless devices (e.g., sensors) with processing/storage resources in their proximity in network topology terms. Indeed, several flavors of this idea, including micro-data centers, cloudlets and Fog Computing itself, follow the Edge Computing model [37], by which data/computations are processed using computing resources located at the edge of the network – accessible through wireless protocols – and optionally using remote resources in the Cloud. As Edge Computing is far more recent, literature reviews are scarce compared to MCC. Particularly, [38] reviews the security treats inherent to Fog Computing environments and delineates potential solutions. The most recent survey found [39] discuss Edge Computing approaches aimed at designing offloading strategies for *mobile* clients, and bases the analysis in three dimensions: offloading decisions, resource allocation and mobility management. Unlike this work, [39] suggests the possibility of using mobile devices as destination for offloading computations/data, but the survey is focused on techniques augmenting mobile clients via fixed computing resources (e.g., local servers and computer clusters).

Table 1 provides a high-level comparison of MCC and Edge/Fog Computing. MCC aims at augmenting mobile clients using Internet-accessible computing resources (higher network latencies), while Edge/Fog Computing brings resources close to the client (lower network latencies). Moreover, Edge/Fog Computing aims at primarily exploiting local resources and optionally Cloud computing nodes, thus introducing the idea of resource hierarchy [39] for handling computations and data. In either paradigms, particularly, the role of SMDs is exploiting external resources to circumvent their inherent limitations.

Moreover, the last row of the Table represents a complementary form of Edge/Fog Computing by which computations/data can be offloaded to nearby SMDs as well. This idea, termed "Dew Computing"[1] by some researchers [40,41], conceives SMDs as having a dual role by which they both exploit nearby and remote fixed resources, and also offer their own resources (CPU cycles, storage and even sensors) to other clients. To the best of our knowledge, there are no surveys published in the topic, and hence this is the first work reviewing techniques focusing on resource allocation using SMD as computing resource providers in Edge Computing. Our review is guided by a characterization of the singularities of SMDs compared to fixed computers, the challenges these singularities pose to addressing resource allocation in Edge Computing environments, and the approaches proposed in this line so far.

3. Preliminary notions

Resource allocation is a central theme in distributed computing environments since it determines the effectiveness with which resources are utilized for accomplishing computational tasks (from now on simply "tasks"). Broadly, resource allocation is described as a decision-making process of allocating resources to tasks. In the literature, this general process have been referred as scheduling, task mapping, matchmaking, task distribution, resource scheduling, resource selection and resource provision. However, along this survey we will refer to it as resource allocation. Moreover, a task, also called job, is an atomic computation whose execution starts and finishes on the same node.

In distributed computing environments like clusters, Grids or Clouds, RA is performed via wired links, because resources are provided by fixed machines connected through wired interfaces. However, when considering arrangements of SMDs such RA process occurs via wireless links. Section 3.1 provides details of the typical forms of arranging SMDs adopted by the state-of-the-art RA proposals. With regard to resources, SMDs provide context sensitive, communication and computing capabilities [42]. Individually, none of these capabilities are exclusively found in SMDs. For instance, wireless communication and sensing capabilities are found in wireless sensors, and many RA mechanisms have been proposed for exploiting such capabilities, specially in the sensor networks research area [30,43]. The same occurs with SMDs computing capabilities, since these are not different from those found in fixed hardware, for which also, many RA mechanisms have been proposed [29].

Then, if SMDs resources are not distinctive and there are many works that propose RA mechanisms for exploiting such resources, an evident question is what makes the process of allocating resources in SMDs networks different from that of allocating resources offered in mobile sensors networks or fixed computers

[1] The term stems from the metaphor of how water condensates at different granularity levels (dew, fog and clouds).

Table 1
Mobile Cloud Computing, Edge/Fog Computing and Dew Computing: Main characteristics.

Paradigm	Resource proximity	Offloading target	SMD role	Related surveys
Mobile Cloud Computing	Low	- Remote Clouds	Exploit external resources	[10,12,32–35]
Edge/Fog Computing	High	- Local servers/clusters - Remote Clouds	Exploit external resources	[38,39]
Dew Computing	High	**- SMDs** - Local servers/clusters - Remote Clouds	Exploit external resources **& contribute with own resources**	**This survey**

Table 2
Reviewed resource allocation mechanisms characterized by networking support.

Networking support	RA proposals
Ad-hoc	[19,44–52]
Infrastructure-based	[53–66]

networks. In short, the reasons are the SMDs singularities derived from their unique mixture of capabilities, constraints and intended purpose. A detailed explanation of these singularities is provided in Section 3.2. More explicitly, SMDs are not single-purpose devices like sensors but multi-purpose devices like PCs. However, they cannot deliver "infinite" resources like PCs because, like sensors, they rely on batteries to operate. But, unlike many kinds of sensors, SMDs batteries support several charging–discharging cycles, which means that, despite their resources are not available continuously over time, these are available for several discrete time periods. Moreover, like PCs, SMDs are used by people, which means that they cannot be considered dedicated devices to purpose-specific tasks anytime, anywhere, as the case of sensors. However, like sensors, SMDs use wireless networks and the communication is feasible even while they are moving, nonetheless, such characteristic diversifies the energy costs of maintaining an active connection and transferring data. Therefore, the dynamic operation and heterogeneity introduced by the above features pose new challenges to the resource allocation problem when exploited resources are provided by SMDs.

3.1. Networking support for coordinating SMDs collaboration

In order to exploit SMDs computing resources, state-of-the-art RA mechanisms utilize a networking support to acquire knowledge of resource availability, distribute tasks and collect results. Fig. 1 illustrates examples of the most common networking supports under which these RA mechanisms operate, differentiating between infrastructure-based and ad-hoc networks. Table 2 complements Fig. 1 with the distribution of RA works reviewed in this survey according to the type of networking support adopted.

RA mechanisms assuming infrastructure-based networks exploit the communication range of dedicated communication hardware, base stations (BSs) like WiFi access points or cellular towers, to communicate with SMDs. A proxy-based setting is a representative example of such kind of RA mechanism, which centralizes SMDs knowledge through a special fixed node, called proxy. The proxy acts as a gateway to high bandwidth backbones or abstracts details of SMDs to higher levels of a hierarchical resource organization [39].

The other type of RA mechanisms operate assuming ad-hoc multi-hop network that shares similarities with P2P networks with respect to self-organization, decentralized control capabilities, and connectivity in highly dynamic environments. Examples of these type of networks are DTNs, OppNets and MANETs, where nodes operate as hosts and routers, i.e., they contribute with resources not only for executing tasks but also for forwarding packets towards other nodes that might not be within the direct transmission range of the source node [26]. Ad-hoc networks have been proposed to support communication where a fixed networking infrastructure is not available, e.g., in rescue operations after natural disasters [67–69].

3.2. SMDs singularities

Accurate resource quantification is a crucial feature for an RA mechanism to be effective. In fact, when resources are very heterogeneous and their availability dynamically varies, the RA process becomes complex. Resources provided by SMDs are very heterogeneous, which results not only from the combination of different hardware models of CPU, GPUs, sensors and wireless radios that manufacturers include in a single device, but also from the modes in which some of those circuits are able to operate [70]. For instance, CPUs voltage and frequency can be dynamically scaled (DVFS) to balance performance and energy consumption.

Besides, another fact that contributes to resource heterogeneity is the finite energy of SMDs. Dedicated devices with the same hardware but different remaining energy are not expected to provide the same amount of resources to execute tasks. Supposing that the amount of resources only depends on the SMD remaining energy, which in turn is difficult to estimate accurately, then a rank of resource quantity could be relatively easy to obtain by ordering SMDs by their remaining energy. However, when the remaining energy is not the only different feature of SMDs, i.e., their hardware characteristics also differ, such resources quantification becomes more complex. This complexity increases even more when considering that resources availability dynamically changes as consequence of the multipurpose and non-dedicated nature of SMDs. Below, we delineate the aspects that contribute to the resource heterogeneity and dynamic availability – from now on called SMDs *singularities* – , which state-of-the-art resource allocation mechanisms acknowledge for the effective exploitation of SMDs:

- *User Mobility*: It refers to the fact that SMDs location depends on their owner mobility. It is a singularity that strongly affects the quantification of SMDs communication resources. The higher delays, error rates and frequent spurious disconnections that wireless communication of SMDs presents when compared to wired communication is, in part, due to user mobility. In ad-hoc networks, mobility can force the path to reach a destination node to change during the data transmission process. This, in turn, alters communication time, energy spent and intermediate nodes involved [71]. In infrastructure-based networks SMDs mobility can increase handoff events, which are an important source of delay in the data transmission process [72]. Even when path changes or handoff events do not occur, the signal quality of an SMD link is altered with the user mobility and that, in turn, affects the rate of useful data that is being transferred. The user mobility may derive in connection–disconnection patterns with regard to, e.g., access points in infrastructure-based networks or other SMDs in ad-hoc networks. In certain type of networks, like DTNs or OppNets, the connection–disconnection patterns, rather than being considered faulty

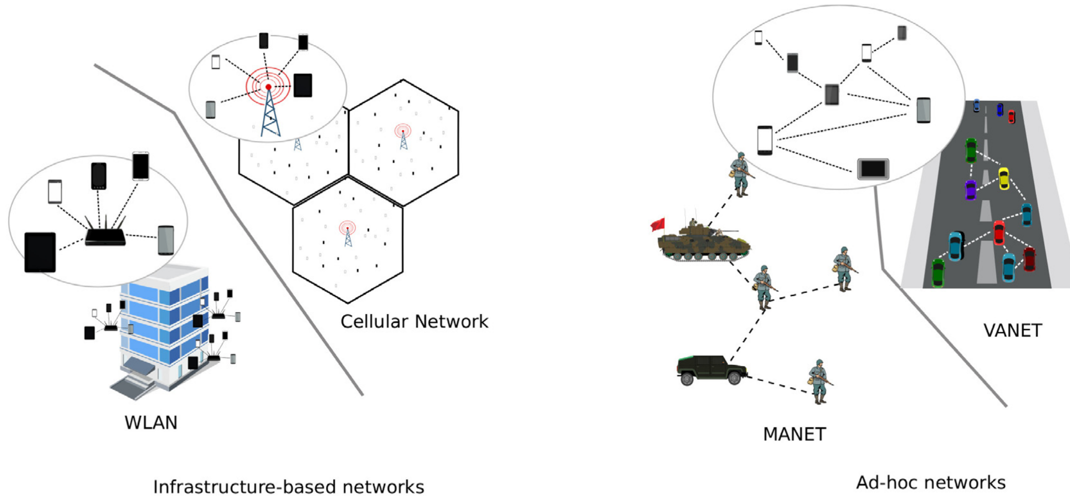


Fig. 1. Networking support for arranging SMDs.

situations, are treated as normal operation conditions under which resources should be exploited [73]. User mobility does not only contribute as a source of heterogeneity when exploiting communication resources, but also when exploiting sensing capabilities of SMDs. As was pointed out in a discussion of a participatory sensing survey [74], the uncontrolled mobility of people is challenging because it leads to the *sensor availability problem*, i.e., the rendezvous among the sensors and the communication infrastructure may not happen on the time scales best suited to the needs of applications.

- *Lack of ownership*: It refers to the non-dedicated nature of SMDs resources, i.e., to the fact that the system or application that aims at scavenging SMDs idle resources does not own the control of resources. In other words, resources such as CPU time, are shared with SMD owner processes and applications. Irrespective of the incentives that encourage owners to offer their SMDs' resources [75–77], external tasks are expected not to (heavily) degrade the performance of owners' applications or experience. Such scavenging rule suggests that the actual resources throughput level and availability differ from the maximum level that the SMDs hardware is able to deliver. This issue is not exclusive to SMDs but for desktop PCs of volunteer computing, e.g., the BOINC and XtremeWeb platforms [78]. However, with SMDs the non-dedicated issue is brought to foreground because they are primary computing devices with which people interact very frequently during the day. For this reason, resource scavenging should be performed in such a way that it does not degrade owner's applications performance to unacceptable QoS levels. Besides, careful energy management must be ensured, i.e., foreseeing future owner interactions.
- *Exhaustible Resources*: It is related to the fact that resources provided by SMDs are limited by the energy availability of their batteries. This poses a new challenge to RA mechanisms since finite energy is a new dimension of heterogeneity at the moment of quantifying the resources that an SMD is able to provide. Resource quantification is a top-priority concern of RA mechanisms [16,79] that target battery-driven devices. Even when resource exploitation is planned to occur while SMDs are plugged to the electricity grid, the scenario is not challenge-free. As shown in [80], such type of resource scavenging should be regulated so as to not heavily affect the time that the charging cycle lasts. Recall from the above singularity (Lack of ownership) that reserving energy for future owners interactions is important since SMDs are non-dedicated devices.

From a practical point of view, the impact of singularities on achieving effective SMD source exploitation may vary from one resource exploitation opportunity to another. For example, is mobility an issue that should be addressed by the RA when the exploitation scenario involves stationary SMDs in a place of interest? Does lack of ownership represents an issue when SMDs to be exploited are in places where owners are supposed to not interact with the SMD very frequently? Is energy a primary concern when SMDs are plugged? In short, a resource exploitation opportunity involves aspects such as a place, a social context, a timescale, a type of communication support and tasks whose completion is performed with a set of SMDs. All these aspects suppose the existence of a wide variety of resource exploitation opportunities where some of the singularities may not be a concern that needs to be fully addressed by the RA mechanism. An example is user mobility when the SMDs involved are exploited at the same place during the same time period, like during office hours [53].

Indeed, the fact that most of the current RA proposals targeting SMDs as resource providers do not simultaneously address all these singularities, does not invalidate their contributions. By contrast, addressing a subset of SMDs singularities relates to the resource exploitation opportunity these proposals target. For this reason, and to avoid focusing simply on whether a work address or not a singularity, this survey delineates a practical categorization of resource exploitation opportunities based on the importance or weight of each singularity on the resource exploitation approach proposed. The categorization serves as a criterion for organizing current and future RA mechanisms targeting SMDs, as well as to better appreciate their contributions and facilitate future comparisons.

4. Resource exploitation opportunities

SMDs have become the primary computing device carried by people most of the time. This phenomenon attracted the attention of many researchers who see in social contexts such as libraries [81], university campuses [82], conference rooms [51], museums [21], or even outdoor public places [83] natural opportunities for scavenging their aggregated computing capabilities. There are also researchers [47,66] who differentiate the computing potential of co-located SMDs in stable, unstable and very unstable, or equivalent coarse-grained categories, based on their connection stability. Other works [49,56], in turn, incorporate the notion of connection predictability. In [56], a stability spectrum is illustrated with real life examples where the least stable and predictable settings are the opportunistic networks [49,51], while the most stable setting is represented by a set of mobile CPUs mounted in single chassis [14]. Between these extremes, are settings where people congregate with its personal devices, e.g., at known time schedules or during certain time periods, such as in a classroom, a theater, a museum, while traveling by bus/train, or stopped at traffic lights, among others.

The computing potential characterization of SMD groups based on their connection stability covers only communication-related capabilities of SMDs, and leaves outside aspects concerning the computing-related capabilities. Given that offloaded tasks to an SMD share resources with that of its owner, his/her interaction should be considered for rating the real capabilities. This involves categorize the computing capabilities by looking at, for instance, the most used applications, frequency and length of sessions. A group of SMDs whose owners present high-demand usage profiles might not render the same computing capability than a group whose owners present low-demand usage profiles. The first could be the case of people entertaining themselves with a resource-intensive game while in a waiting room, e.g., at the doctor office, an airport, etc. The case of low-demand user profiles may be the case of students attending motivating lessons in classrooms of an university campus. In the middle of interaction categorization, there are clients at a coffee shop, surfing the web or reading an electronic newspaper.

The heterogeneity that SMDs singularities introduce to resource quantification should not be analyzed only in terms of social contexts dynamics. The networking support used to coordinate resources also affects how singularity challenges are presented. To illustrate the idea, suppose that it is desirable to employ SMDs computing resources by exploiting the regularity of people driving their cars through a highway. Besides, consider that resources coordination is either through a set of cellular towers placed along the highway (infrastructure-based support) or through VANETs (vehicular ad-hoc networks). With the first coordination support, by considering SMDs location information, delays in responses can be mitigated by avoiding schedule data transfers to SMDs traversing areas with high communication latency, e.g., where handover operations are expected to occur [84–86]. With the second coordination support, the fact that clusters of SMDs move in the same direction at a similar speed do not negatively affect the established links, and the treatments of mobility issues are designed for special cases rather than for normal operation modes.

To acknowledge how complex would be to provide a categorization for measuring the computing potential of SMD clusters, we do not have to forget the characteristics of the tasks [46] that an SMD cluster could be in charge of execute. In concrete, a bag of independent compute-intensive tasks (e.g., number crunching) might be less affected by disconnections caused by user mobility than compute-intensive tasks with data dependencies, e.g., a workflow. While in the first case, the collection of partial results only would be delayed by interruptions in the communication, in the second case, the entire application execution can be delayed, since tasks might not be able to start until the data produced by all preceding tasks is available. Similarly, tasks with hard deadlines will be more affected by high-demand usage profiles than tasks with soft deadlines.

All in all, identifying which singularity/ies need to be addressed on every possible resource exploitation opportunity would be impracticable due to the innumerable possibilities of combining SMDs usage contexts, networking supports and task characteristics. Alternatively, groups of resource exploitation opportunities can be associated to heterogeneity posed by different singularities combination. Table 3 outlines the distribution of state-of-the-art RA proposals for different singularities combinations. *UM* refers to user mobility, *LO* to lack of ownership and *ER* to exhaustible resources. The $\sim$ symbol on top a singularity acronym is used to indicate that the heterogeneity type posed by the singularity does not characterize the group of resource exploitation opportunities, or equivalently, it is not a concern that affects the effective exploitation of SMD resources. Then, works placed in a group with one or more singularities decorated with the aforementioned symbol, should be interpreted as RA works not able to deal with, or at least, not evaluated for, the resource heterogeneity type posed by the singularity. Notice that the resource exploitation opportunity group where none support for singularities is needed – the $\widetilde{LO}\ \widetilde{ER}\ \widetilde{UM}$ combination –, is marked with N/A. It indicates that outfitting RA mechanisms with special logic to cope with SMDs singularities would not be necessary for effectively exploiting resources. Moreover, cells filled with the cross symbol (✕) indicate that there are no RA works in the literature addressing such singularities combinations.

In order to consider an RA proposal as qualified for a type of resource exploitation opportunity, it should address the heterogeneity imposed by the singularities combination. Addressing a singularity means that specific information, models or parameters related to the singularity are used to propose and/or evaluate a theoretical, or practical, resource allocation algorithm. Examples of this include the application of conditional probabilities and Markov chains to model connection/disconnection of SMDs within a region of interest, the usage of GPS location data and remaining battery to build rankings of SMDs with which weight their reliable/power of computing capabilities, among others.

As said, a singularity can be addressed from the evaluation perspective. For instance, there are proposals that do not contemplate concrete actions to deal with user mobility, but the performance evaluation comprises settings where nodes are configured to change their location or enter/exit the operation range of the RA mechanism [19,50,51,60,87].

The next subsections provide details of the surveyed RA mechanisms addressing each singularities combination. It worth noting that the description provided does not aim to be a full characterization of each work but a highlight of techniques, algorithms and information related to the singularities combination they address. Common features of the surveyed works, that complement these descriptions can be found in the analysis presented in Section 5.

4.1. RA proposals addressing heterogeneity derived from UM_ER singularities

In [49], the authors provide concrete actions for dealing with user mobility and resource exhaustion by means of Serendipity, a system for disseminating tasks in a group of SMDs intermittently connected through ad-hoc links. They propose three tasks dissemination strategies based on the predictability of future contacts and the existence of a control channel for coordination. When future contacts are predictable and there is a control channel (ideal case), it is proposed a Water-filling greedy strategy that iteratively chooses the destination SMD for every task that pursues global minimization of tasks completion time. When future contacts are predictable but there is not a control channel for coordination, a Computing on Dissemination (CoD) strategy is used to distribute tasks opportunistically and tasks time minimization is performed locally. For the case with least context information, i.e., unpredictable future contacts and no control channel, the authors proposed upCoD, a variation of the CoD strategy. Instantiation of the strategies are also proposed for increasing the lifetime of participant SMDs where the residual energy is considered for tasks

Table 3
Resource exploitation opportunities resulting from a binary valuation of addressing SMDs singularities.

	$\tilde{LO}$		LO		
	Ad-hoc	Infrastructure-based	Ad-hoc	Infrastructure-based	
ER	[19,44,47,49,50]	[54,60,62]	×	×	UM
$\tilde{ER}$	N/A		×	×	$\tilde{UM}$
$\tilde{ER}$	[46,48,81,87]	[56,58]	×	[64]	UM
ER	[45,88]	[59,61,65]	×	[53,55,57]	$\tilde{UM}$

dissemination. Tasks are assumed to be provided with information that allow the Serendipity system to determine the execution time and energy spent in each candidate SMD. The evaluation includes a set of experiments with a simulation software using real traces and another set with a prototype implementation of Serendipity that uses real SMDs, a speech-to-text application and a face detection application.

In [47], Abderrahmen Mtibaa et al. propose a resource sharing algorithm for ad-hoc networks where nodes collaborate on the execution of independent computational tasks. The objective of the algorithm is to increase network lifetime, i.e., prolong the time of the first node that fails due to a battery depletion event. The algorithm considers finite energy of nodes and intermittent connection when offloading tasks among nodes at one hop, and two-hop distance from the node that initiates the offloading process. For each task to be executed, the algorithm evaluates the energetic convenience of delegating it to a neighbor node versus executing the task locally. The criterion to define such convenience uses the communication and computation requirements of tasks, the energy spent by nodes while executing and transferring the task and the remaining energy of nodes. For dealing with intermittent connections and before evaluating the energetic convenience, a check for the existence of communication paths is performed. The paths should exist to allow the tasks distribution and results collection to happen within the tasks deadlines. The authors assume that the group of collaborating nodes can be derived from the analysis of contact duration and frequency information. Once such group is identified, the algorithm continues with the evaluation of the energetic convenience. The proposal is evaluated with experiments that include real mobility traces and energy consumption information derived from real mobile devices.

In [60] the authors propose a service composition mechanism for negotiating SMDs resources using agents. The focus of the proposal is on an economic model that rules the requests and offers of resources, which aims at maximizing the utility of a resource allocation by the application of a Lagrange multiplier-based approach. The resource exhaustion is contemplated by the constraint of SMDs finite capacity in the utility maximization problem statement. The user mobility singularity is not considered in the approach other than through its evaluation. The authors include simulation scenarios where SMDs move with a random-walking mobility pattern with speeds varying between 1 and 20 meters per second.

In [50], Furthmüller et al. propose a framework for sharing resources of SMDs. The usage of a resource or group of resources are offered and consumed by applications as services. A study of several service selection criteria is presented with focus on determining the benefits on extending the battery life of service providers. The proposed selection criteria use information of service energy consumption and the remaining charge of devices. Additionally, a framework to derive the energy consumption model of SMDs is proposed. Moreover, user mobility is addressed in the same way as [60], i.e., only through the evaluation of the approach.

In [62], the authors propose and simulate the performance of an algorithm that predicts SMDs availability by using a physical availability factor (PAF) and a battery availability factor (BAF) for scheduling heterogeneous tasks. The authors assumed that SMDs providing resources are connected to a base station that is part of the infrastructure of a cellular network. Six adjacent cells of that network are considered a zone. The resource allocation is structured in a two-level hierarchy where the first level operates from the central cell of a zone and hides resources heterogeneity details to the inter-zone level. The PAF is used to infer SMDs future location and is determined based on its movement type, i.e., moving towards and moving apart from the base station, and the mobility pattern. The movement type is predicted based on Markov chains, SMDs GPS information and the Haversine formula. The latter is used to calculate the distance, in this case between an SMD and a reference base station, taking into account the curvature of the earth. Moreover, the BAF value is determined with parameters such as SMDs battery capacity, C-rate, battery power usage and total battery availability. Authors propose to use both factors as criterion for measuring nodes availability.

In [19] the authors propose an RA scheme appropriate for local mobile clouds that assigns tasks to energy efficient processing nodes with an adaptive probabilistic approach. The proposal is not aware of exhaustible resources, but the singularity is lightly contemplated in the evaluation when authors state that SMDs are initialized with the same battery level. The RA mechanism operates as follows: for every task, it selects a set of SMDs able to execute the task within a time constraint determined with information of SMDs processing capabilities. The definitive SMD in charge of the task execution is probabilistically selected from the set of candidates where probability is defined based on the energy each SMD spent on executing the task. The time constraint is, in part, defined by the task deadline, which is provided by the task owner, minus a margin value that is calibrated when tasks fail. Task failures are detected when tasks cannot be completed within the owner task deadline and one cause of these failures is attributed to unpredictable SMDs queuing delays which are in turn consequence of assignments from multiple source nodes. The RA mechanism is evaluated in a simulated environment with stationary and in-movement SMDs settings, varying the number of SMDs, their computing capabilities, mobility patterns and tasks requirements. Other variable taken into account is the interval of topology control messages in the task completion rate.

In [54] the authors present a framework that combine concepts of quantum computing and binary gravitational search to process jobs with execution dependencies. The resulting meta-heuristic (QBGSA) aims at minimizing the turnaround time of jobs. The exhaustible resources and user mobility singularities are considered from the fitness function used to evaluate the solutions quality found in each iteration of the evolutionary process. Precisely, the fitness function combines a *BP* (battery power) value and a *MS* (mobility score) value. The first is derived from the remaining energy percentage reported by SMDs, and the second is calculated using information of the mobility history of SMDs through different coverage ranges during a week for office time hours. The performance of QBGSA is compared with that of QGA (quantum genetic algorithm).

In [44], a resource provisioning framework with autonomic capabilities is proposed for scheduling independent tasks of real-time in-the-field healthcare applications. The devices are assigned with roles that relate to their capabilities for computing, sensing and transferring data. Nodes with the service requester (SR) role ask for some data to be processed. Nodes playing the service provider (SP) role sensor data from the environment and/or offer computing capabilities while a node with the role of arbitrator (broker) provides resource discovery and allocation services. The RA logic leverages long-term statistics as a way of managing uncertainty caused by users mobility. Statistical indicators considered are the average arrival rate of SP nodes and the average times SPs are connected and disconnected from an arbitrator. Besides, the novel concept of *applications waypoints* outfits the RA logic. It is an aid to mitigate the uncertainty of tasks execution performance caused by unpredictable SPs resource utilization, inaccurate energy estimation and tasks completion time. The RA can be tuned with policies to achieve a minimization of the maximum battery drain – lifetime maximization of the network –, or minimization of the applications response time without considering battery drain QoS maximization —. The policies operate with information of nodes availability periods as well as performance metrics such as energy consumption and resources utilization. Provider nodes voluntarily inform such information to arbitrator nodes.

4.2. RA proposals addressing heterogeneity derived from LO_UM singularities

In [64], the authors provide some insights on the impact of lack of ownership and user mobility singularities in the completion of CPU-intensive tasks. The work proposes an economic model for resource allocation based on non-cooperative bargaining game theory. In addition to that, they assess the impact of reducing the cost of location update (mobility tracking) of SMDs in an IEEE 802.11 mobile Grid architecture by proposing a location management framework based on the Lempel–Ziv (L778) compression algorithm. They show that using the data provided by such framework, the task allocator component, accommodates more tasks than when such data is not considered. Besides, the work contemplates the non-dedicated nature of SMDs by considering non-grid tasks competing for resources. The execution model of devices is assumed to be a M/G/1 preemptive priority queue where internal tasks arrival rate can preempt grid tasks.

4.3. RA proposals addressing heterogeneity derived from ER_LO singularities

In [57], the authors boost the SEAS [89] mobile Grid scheduler with different job stealing techniques. The way SEAS consider the exhaustible resources of nodes is by predicting remaining node uptime from battery drop events. Remaining uptime information combined with FLOPS and assigned tasks is used to rank SMDs and decide the most appropriate one for executing a newly arrived CPU-bound task. The job stealing techniques are proposed as a protection mechanism against unbalanced load produced as consequence of inaccurate remaining uptime predictions that, in turn, derive from factors such as workload, network usage and inaccurate battery sensor of SMDs. Experiments include settings with dedicated and non-dedicated SMDs. In non-dedicated settings, SMDs CPU are configured to have 30% utilization simulating an average owner CPU utilization.

In [55], the authors explore several RA mechanisms that use novel and practical criteria for ranking SMDs. The criteria includes an enhancement of SEAS [89] that simplifies the way exhaustible resources are considered, and other alternative criteria derived from benchmarks, manufacturer battery information and historical tasks completion time. The SEAS enhancement consist in the usage of battery state of charge value instead of predicted remaining uptime in the SMD ranking formula. The change is proposed to mitigate a cold start effect in the ranking formula. In particular, the cold start affects the prediction of the remaining uptime, which requires several samples of an SMD state of charge. Until these samples are available, devices – specifically smartphones and tablets – are not accurately ranked causing unbalanced task distribution. An alternative criterion for ranking SMDs is JEC which replaces the FLOPS component of the E-SEAS ranking formula by a job energy consumption rate that combines benchmarking information and battery capacity of SMDs provided by manufacturers. The third alternative criterion called FWC considers historical execution time of tasks to infer SMDs future performance. All criteria use SMD state of charge updates to contemplate the resources exhaustion singularity. Lack of ownership is considered via experiments with settings of dedicated and non-dedicated SMDs in the same way as [57].

In [53], a two-phase scheduling approach that uses the SMDs rank criteria presented in [55] is proposed. The two-phase approach aims at mitigating unbalanced task distribution caused from tasks heterogeneity, out of date/inaccurate estimation of remaining energy and lack of owner interaction control influencing the availability of CPU cycles. In a *first scheduling phase* , CPU-intensive jobs are scheduled in a centralized fashion as soon as they arrive to the scheduling component, by using practical criteria to rank SMDs [55]. The second scheduling phase redistributes load to subexploited SMDs employing decentralized job stealing techniques where SMDs play an active role in scheduling decisions.

4.4. RA proposals addressing heterogeneity derived from UM singularity

In [48], Shah et al. propose the Two-Phase Resource Allocation scheme (TPRA) for dealing with the user mobility singularity. TPRA exploits movements history information of SMDs and wireless communication energy consumption properties to distribute dependent tasks of varying requirements. The first phase aims at reducing the probability of communication interruptions due to SMDs mobility. When an SMD needs computing resources, it broadcasts its next probable location and SMDs willing to provide such resources in that location reply to the broadcast message. The next probable location is determined by means of a Markov chain that stores the history of user mobility patterns. The physical area where nodes movement is mapped to a virtual grid equal-sized cells. The states of the Markov chain are the cells while transitions from one state to other represent the movements of an SMD between cells. Movements of each SMD between cells is represented as a probability matrix whose values are updated every time an SMD moves. The second phase of the RA scheme uses physical distance information among SMDs to minimize communication latency of tasks that need to exchange data. In [46], an extension of [48] called ERRA is proposed, which aims at minimizing SMDs energy consumption while executing a group of tasks. In ERRA's second phase, data dependent tasks are assigned to the heaviest weight k-devices group, where k is the number of tasks and the weight of groups is determined with information of the SMDs transmission power level.

In [58] the authors propose a fault avoidance approach for scheduling tasks by taking into account dynamic properties related to communication capabilities of mobile devices. Such properties are: availability, reliability and maintainability – whose individual values are aggregated into another single property called effectiveness –, and d-effectiveness – that quantifies users' movements pattern –. Availability is defined as the probability that a device is operational and/or able to return results which, in turn, depends on the time period the device is able to provide its resources. Reliability is the probability that a device performs a task for a given time period without failures and maintainability is how quickly a device recovers from a failure. The product of these properties values results in an *effectiveness value* which is used to rank devices. Device *A* with higher effectiveness value than device *B* is supposed to be available for a longer period to perform tasks, to operate with less failure probability and to recover quicker when a failure occurs than device *B*. D-effectiveness property is the ratio of effectiveness to the Euclidean distance between the availability, reliability and maintainability values in the present with regard to the same value measured in the past. The prediction quality of d-effectiveness depends on the fact that users exhibit regular connection patterns w.r.t the same time-of-day and day-of-week. The evaluation includes the usage of real-life traces from SMDs.

In [56], Habak et al. propose FemtoCloud, a system for scheduling independent hybrid tasks, i.e., with data and CPU cycles requirements. A prototype of the system was implemented for Android-powered SMDs. The proposal aims at exploiting the computing capabilities of a group of SMDs putting special attention in the connection periods of SMDs to *the controller*, a component in charge of the task assignments. The time at which tasks should be assigned and results collected are scheduled by two complementary greedy heuristics that take into account the tasks deadline constraints. One heuristic schedules the time and resource utilization for transferring inputs and executing the tasks. Another heuristic schedules the appropriate time for SMDs to send the tasks results. The objective is to achieve the highest amount of useful computation done. The heuristics efficiency is evaluated through simulations and real arrangements of SMDs.

In [87] it is evaluated the performance of a set of classical heuristics and a set of adapted heuristics inspired on the classical ones for scheduling jobs in ad-hoc networks comprising SMDs. The evaluation measures tasks/user centric parameters including average makespan, average offloading time, average waiting time and average tasks slowdown, and system centric parameters, i.e., average resource utilization. The classical or "seminal" heuristics considered in the study are the MET, MCT, MinMin, MaxMin and Sufferage, while the adaptations which aim at considering the multi-hop communication and delay affecting jobs and results transferring are called MinHop, MetComm, MCTComm, MinMinComm, MaxMinComm and SufferageComm. The proposed heuristics are not designed to deal with or take advantage of future nodes mobility, however, the experimental setup, which is made through simulation, considers nodes mobility according to a random way point model in an area of one square kilometer.

4.5. RA proposals addressing heterogeneity derived from ER singularity

In [59], a variant of the economic model proposed in [64] is presented. It does not consider user mobility related issues, neither SMDs lack of ownership but contemplates resource exhaustion when distributing tasks, a singularity that is not addressed in [64]. In the process of resource negotiation, a resource brokering agency that centralizes the information of SMDs, assigns them a *reliability score*. The score is calculated combining information of tasks requirements and SMDs resources capabilities including processing rate, memory, bandwidth and battery power. The experiments include a comparison of the proposed RA mechanism against others of the same authors (Ghosh et al. [64]).

A multi-objective two-step resource allocation approach that deals with the resource exhaustion of SMDs is proposed in [45]. The first step employs a Non-dominated Sorting Genetic Algorithm II (NSGA-II) to obtain a set of solutions near to the Pareto-optimal front. The second step employs the entropy weight and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to select the best balanced solution that minimizes tasks completion time and consumed energy. Other constraints considered in the proposal are tasks user acceptance deadlines, nodes residual energy and budget. The budget constraint corresponds to a virtual payment for offered resources that is proposed as an incentive mechanism for encouraging SMDs collaboration. The proposal is evaluated with OMMC, a context-aware offloading middle-ware, that was installed on real SMDs. The experimental scenarios include different combination of SMDs, heterogeneous tasks and comparisons with proposals of other authors [49].

In [65], an RA mechanism, named HPSM, targets dependent tasks and the objective function used to guide node selection, can be tuned to prioritize the conservation of SMDs energy, to minimize tasks schedule length or to achieve a balance between both opposite goals. As HPSM is designed for dependent tasks, HPSM considers the execution capability of SMDs and communication bandwidth between them, as well as energy spent during task execution and results transferring. The RA mechanism has a compile-time phase where tasks dependencies, represented as a DAG (Directed Acyclic Graph), are analyzed to determine groups of tasks that can be executed in parallel. Then, tasks are assigned with a priority based on amount of input data they require, the output data they generate, the sum of both amounts of data and the average computation time of the task on every candidate SMD. After that, in a node selection stage, by iterating the list of prioritized tasks, the SMD that minimizes the configured objective function is assigned with the current task. The cycle is repeated until all tasks are assigned. A component of the objective function contemplates the exhaustible resources singularity of SMDs. The performance is evaluated through simulation and comparisons against two algorithms named HEFTM and PETSM that are adapted from HEFT [90] and PETS [91] algorithms inherited from traditional heterogeneous computing environments, varying the number of DAG instances generated, number of SMDs and energy consumption rates of SMDs.

In [88], an RA mechanism for prolonging network lifetime in a collaborative ad-hoc network composed by SMDs is proposed. The tasks are generated by SMDs that also participate in the execution of other tasks. The RA decisions are decentralized and hierarchically organized in clusters. It means that, a task assignment is firstly evaluated among the SMDs within the scope of the local cluster where the task execution request is originated. If none of the SMDs of the local cluster is able to execute it without depleting its battery, the assignment is delegated to an inter-cluster level where SMDs of other clusters are considered. To compute the feasibility of an assignment, the RA mechanism considers the energy required to compute and transfer the involved task. The evaluation is performed with a system prototype where real SMDs and Android emulators are involved.

In [81], it is evaluated the effectiveness of applying Work Stealing techniques as decentralized RA mechanism for distributing resource-intensive tasks in a cluster of SMDs connected through an ad-hoc network. The experiments are designed to measure how different parameters, including topology layout, task sizes and minimum amount of work units exchanged between nodes influence the speedup achieved in the completion of a task. As a case study, the authors use the distributed Mandelbrot set generation to represent the resource-intensive tasks and a cluster of real SMDs connected via Bluetooth. By looking at the way in which Work Stealing technique is applied, one cannot say that the RA is aware of the exhaustible resources of SMDs. However, the fact that the proposal is evaluated with real devices, i.e., under the effects of real battery depletion, makes it qualify as a proposal that addresses the exhaustible resources singularity.

In [61], the implementation of the Ant colony metaheuristic is proposed for scheduling applications in an Hybrid Local Mobile Cloud Model (HLMCM), that is presented as an extension of the Cloudlet architecture where execution resources are provided by a cluster of SMDs forming an ad-hoc network. Applications have varying resource needs that are satisfied with those available at the HLMCM. The completion of each application is associated with an utility value for the system and is achieved by consuming varying resources quantities, available at nodes in finite amounts. The latter is a constraint denoting the way the RA mechanism addresses the exhaustible nature of SMDs resources. The RA problem is seen as an optimization problem that pursues the goal of maximizing the total utility of the system under the finite amounts of available resources provided by SMDs.

5. Analysis of how SMDs singularities are addressed

This section analyzes and classifies how SMDs singularities are addressed by the reviewed RA mechanisms. The analysis and classification emerge from the works that have been described in Section 4, and additional ones whose detailed description have been omitted due to their similarity with such works w.r.t how they address the singularities, their RA goals and the assumptions they made.

Table 4 gives a big picture of RA proposals that are in line with the SMD singularities explained in Section 3.2. One reason that makes an RA proposal to be considered in line with an SMD singularity is that it utilizes information in the RA logic to deal with challenges posed by the singularity, e.g., the current energy level of a device is frequently used to deal with resource exhaustion singularity. The other aspect that makes an RA proposal to consider a singularity is when it represents/includes the singularity within the experimental variables of the performance evaluation. In the first case, the RA proposal is *aware* of a singularity, while in the second case, it is *evaluated* considering a singularity. Furthermore, works that are aware of a singularity are also evaluated w.r.t. that singularity, but the other way around does not necessarily hold.

5.1. User mobility

As shown in Table 5, works that are aware of user mobility are, in turn, differentiated in the way they perceive the singularity. The fact that works such as [46,47,49] perceive user mobility as beneficial, means that user movements are exploited as opportunities for offloading computations and returning results back to the source node. These type of works typically target opportunistic connectivity scenarios [92], i.e., where SMDs contacts occur from time to time. The focus of such RA mechanisms is on searching for time-variant routing paths between SMDs, with the aim of increasing tasks computing parallelism, and in this way reduce tasks completion time and/or balance energy spent by collaborating SMDs.

In contrast, works perceiving user mobility as harmful for effective resource exploitation are [54,56,58,62–64,66]. The fact that nodes move away from the communication range of the component that performs the resource allocation and where results are forwarded, are seen as an important source of tasks failure. To deal with this issue, RA mechanisms employ criteria to measure SMDs service availability, mainly based on historical frequency of connection and disconnection events, length of connection sessions and/or movement patterns inferred using GPS information. A missing feature of practical approaches [58,62,66] is that the criteria used to differentiate SMDs do not consider hardware computing power, which limits the applicability to groups of homogeneous SMDs.

Finally, there are works [19,50,51,60,87] that do not propose concrete actions to exploit user movement but evaluate the performance of the proposed RA mechanisms under user mobility. The schemes used by these works – and those aware of the singularity – to implement user mobility are outlined in Table 6.

User mobility is implemented through simulation by following a synthetic or a trace-based scheme. One type of synthetic scheme uses movements derived from an artificial mobility model, where trajectory shape – circular, linear, random –, number of stops, time between stops and speed range of an SMD are common parameters used to characterize the model. SMDs are then configured with a mobility model, an initial location within a 2D plane and a communication radio. The plane can be associated to the coverage range of some reference node where the resource allocator logic resides [60,62] or to an area where SMDs establish ad-hoc links with other SMDs in their vicinities [19,49,50,87]. The other type of synthetic scheme adopted by RA works for implementing user mobility is by representing only connectivity information of SMDs, i.e., presence/absence of SMDs within regions that are supposed to be from where the RA logic operates. The connectivity information can be represented by probability scores associated to events of connection and disconnection of SMDs [54,63,66], or through in/out time intervals [44,56].

The trace-based implementation scheme utilizes data derived from real SMDs movement. The works that adopt this scheme are [47,49,51,58]. Traces provide data of SMDs connectivity frequency and duration. When connectivity is with regard to other SMDs, for example due to people social interactions, the traces are called *encounter* traces. Popular encounter traces were those originated in the Haggle project, the MIT reality mining project and IEEE Infocom 2006 conference. These, and many others traces are available from the CRAWDAD website [93]. There are also individual efforts [49,81] that built custom encounter traces from the movement study of selected groups of people. When connections are to access points, the traces are called WLAN traces and represent the SMDs user movements through different APs of public or private buildings or campuses [94]. An example of this type of traces is that of the Dartmouth university campus [95] also available from the CRAWDAD website.

5.2. Lack of ownership

The lack of ownership singularity, understood as the sharable condition of SMDs resources by different task submitters, including the SMD owner, has received very little attention by researchers yet. Resources such as CPU cycles for CPU-intensive tasks or network bandwidth for data-intensive tasks should not be assumed as dedicated to perform external tasks but shared with owner processes and applications. One reason for not making such assumption is to avoid overestimating SMDs computing capability. Provided that SMDs owner processes and applications have higher priority over resources usage than external tasks, the load that the former introduce represents a source of heterogeneity that limits the real resources availability to perform external tasks. By ignoring the real resources availability, the completion time/rate of external tasks can be underrated. The other reason is to minimize the external tasks invasiveness on SMDs owners' experience. This means preventing external tasks from overloading an SMD and causing its battery to be depleted without care of the owner usage context, i.e., their future needs, the existence of power plugs to charge the SMD, and/or willingness to share resources on specific situations. Advocating to a context-aware RA might integrate ideas from anticipatory mobile computing [96] where SMDs sensing capabilities are used to infer current SMD owner context, predict his/her future context and/or events and propose a framework to advice the user on future decisions. Table 7 outlines the RA mechanisms aware of and evaluated for the singularity.

Table 4
SMDs singularities addressed by state-of-the-art RA mechanisms.

Singularity	Aware		Evaluated	
	Ad-hoc	Infrastructure-based	Ad-hoc	Infrastructure-based
User mobility	[44,46,47,49]	[54,56,58,62,63,66]	[19,50,51,81,87]	[60]
Lack of ownership	×	[64]	×	[53,55,57]
Exhaustible resources	[44–47,49,50,88]	[53–55,57,59–62,65,89]	[81]	×

Table 5
Approaches for addressing user mobility.

Approach	Perception of user movement	Focus	RA works
Aware	Beneficial (it is an opportunity to reduce tasks completion time)	Search for routing paths through SMDs or potential clusters which favor tasks distribution and results collection	[46,47,49]
	Harmful (it is the main cause of task failure)	Search for SMDs with good connectivity metrics in respect to the resource allocator coverage region	[44,54,56,58,62–64,66]
Evaluated		Add realism to the evaluation of the proposal	[19,50,51,60,87]

Table 6
User mobility implementation schemes.

Scheme	Description	RA works
Synthetic	Movement guided by a mobility model	[19,49,50,60,62,87]
	Movement represented through presence/absence information	[44,54,56,63,66]
Trace-based	Movement information derives from real SMDs location changes (Encounter traces, WLAN traces)	[47,49,51,58]

Table 7
Approaches for addressing lack of ownership.

Approach	Focus	Objective	RA works
Aware	Improve external tasks performance	Exploiting resources to reduce external tasks failure rates (or increase tasks completion rates) in presence of resource fluctuation	[64]
Evaluated		Add realism to the evaluation of the proposal	[53,55,57]

There is just one work [64] with focus only on external tasks performance, which considers resource fluctuation due to user load. The proposal is fed with the rate at which resources are consumed/utilized by all tasks running in every candidate node. Task execution is modeled as a preemptive priority queue where external tasks compete for computing and bandwidth resources with other $P-1$ classes of tasks that represent owner processes and applications. From each class of task, the RA mechanism is fed with task priority, arrival rate and load. This information is utilized to estimate the queue delay of external tasks on every node. Given that such information is highly volatile, the authors plan to study the impact of the refreshing period including network delay parameters.

As Table 7 shows, there are efforts [53,55,57] that, despite not proposing specific criteria to deal with resources heterogeneity associated to the singularity here analyzed, include a preliminary study of the effect of SMDs owner load on tasks completion rate. In other words, these efforts consider the influence of the singularity when evaluating the proposed RA mechanisms. The methodology used by these works and [64] to implement lack of ownership features is through simulation, where two schemes were identified (see Table 8).

In [64], synthetic simulation is adopted, where SMDs processing capability dedicated to external tasks is modeled as a portion of the highest processing capability. Such portion does not vary in time, i.e., fluctuation of CPU or bandwidth while an external task is executing is not modeled. Different SMDs are associated with different dedicated portions.

In [55,57], a trace-based scheme to simulate non-dedicated computing resources of SMDs is utilized. The available computing power of each SMD is modeled through a non-free, or baseline, CPU usage profile. The samples of the non-free CPU usage profile represent the CPU usage changes that indicate a node processing capability occupied by owner processes and applications, i.e., non-available for external tasks execution. In scenarios where dedicated processing capability is simulated, the samples carry information of the CPU usage changes derived from the OS execution and fluctuates around [2–10]% depending on the SMD model and OS version. In non-dedicated scenarios, the CPU usage fluctuates around 30% representing the CPU usage of an average user who performs activities such as surfing the Web, checking emails, using a chat application, among other non-intensive CPU tasks. The fluctuation is not randomly generated but profiled with an Android application which runs in background, and can be configured to hold the CPU usage in a preset level. In [53], the information carried by a CPU usage trace was improved. The concept of user sessions with length and time between sessions parameters were included to reflect a more realistic CPU usage derived from owner interaction. These parameters were derived from empirical user interaction models of a third-party study [97].

5.3. Exhaustible resources

Table 9 outlines the RA mechanisms that address exhaustible resources. The works aware of the singularity are focused either on a conservative or a greedy resource exhaustion policy. Those in the first group aim at extending, as much as possible, the time of the first SMD leaving the network due to battery depletion. In those works, the concepts of fairness [98] and/or network lifetime are commonly used to measure the quality of RA decisions. Works that promote a conservative resource exhaustion can be, in turn, pure conservative (when the extension of the network lifetime has top-priority over other criteria used in the RA decisions) or

Table 8
Lack of ownership implementation schemes.

Scheme	Description	RA works
Synthetic	Non-available resources simulated via arrival rate of SMDs owner tasks	[64]
Trace-based	Non-available resources simulated via CPU profiles	[53,55,57]

Table 9
Approaches for addressing exhaustible resources.

Approach	Focus		RA works
Aware	Conservative resource exhaustion	Pure	[50,88]
		Light	[44,45,47,49]
	Greedy resource exhaustion		[53–55,57,59–61,89]
Evaluated			[19,81]

light conservative (when other than the aforementioned criterion, e.g., completing tasks before the deadline is prioritized).

An example of pure conservative RA mechanism is [88] where a bottom-up strategy guided by nodes residual energy is used for assigning tasks. The scheduling logic is hierarchically organized in a set of interconnected cluster heads, each one in charge of the local scheduling decisions within their vicinity. The head of the nodes cluster from where a task is submitted tries to schedule the task, locally first, within SMDs of the cluster. If no SMD of the local cluster is able to execute the task with its available energy, the task scheduling is delegated up to an inter-cluster scheduling level where the task execution is evaluated by the other cluster heads of the network. The less affected cluster – in terms of energy spent – by the execution of the task is assigned with the task. Other examples of pure conservative approaches are the service selection strategies called "Fair spending" and "Remaining charge" presented in [50]. The Fair spending strategy distributes tasks (service requests) in such a way that each SMD spends the same amount of energy, while the Remaining charge strategy selects the SMD whose remaining charge is less affected by the task execution.

By contrast, light conservative RA mechanisms do not have energy conservation as the top-priority objective but also pursue a QoS-related objective, e.g., violating as few tasks deadlines as possible. Examples of these RA mechanisms are [44,45], which propose multi-objective optimization approaches that aim at balancing tasks completion within tasks deadlines and maximization of the minimum residual energy of SMDs. In [47], tasks energy consumption is evaluated on every candidate SMD at one and two hops distance from the node willing to offload a task. The offloading cost is compared to the local execution and the option that maximizes the residual energy and allow the task to be completed within its deadline is selected. Similarly, the energy-aware offloading version of the heuristics presented in [49] target the minimization of an utility function that contemplates tasks energy consumption on every candidate node and nodes residual energy without ignoring task deadlines. Notice that in these works, the extension of SMDs lifetime is considered upon every task assignment, provided this does not compromise the tasks execution deadlines.

On the other hand, greedy RA mechanisms adopt a resource exhaustion that is not focused on making the whole distributed system last longer, but in exploiting resources to achieve the highest profit, the highest throughput, or the lowest tasks completion times. In those cases, RA mechanisms are aware of SMDs resource availability at the time of evaluating the feasibility of RA decisions. Examples of this type of proposals are [59,60] where resources utilization is ruled by a market-based model where SMDs resources are offered at prices that resource consumers pay to the distributed system in exchange of certain QoS. By taking into account the resource providers energy constraints, the RA decisions must assure the QoS expected by the resource consumer while maximizing the profit. Other examples which follow a greedy resource exhaustion are [53–55,57,89]: they predict the resource quantification using information of remaining charge reported by the OS API to decide the number of tasks each resource provider should be assigned with. In [61], resource exhaustion is not inferred from SMDs remaining energy. The RA is modeled as a multidimensional 0–1 knapsack problem, where each SMD provides fixed countable units of each resource type. For instance, resource provider p provides u units of resource type r. Hence, given the available amounts of each type of resource each SMD provides, the amount of resources each task requires, and the profit each task generates, the RA decisions are focused on maximizing the global profits of the distributed system, subject to the resources constraints of resource providers.

RA mechanisms prioritizing the utilization of energy-efficient resources, e.g., the "Minimum energy" strategy of [50], or ERRA [46], do not qualify neither as conservative nor greedy resource exhaustion approaches because RA decisions are not evaluated based on the current and/or future impact on residual energy. This causes that the most energy-efficient providers leave the system very early compared to less-efficient ones.

Table 10 summarizes aspects related to the implementation of SMDs exhaustible resources. One of the aspects is the methodology employed to account for resource exhaustion, while the other aspect relates to whether the methodology considers some sort of baseline energy consumption, i.e., other than those specifically caused by tasks execution, e.g., energy consumed by SMD owner interaction, OS runtime, connectivity maintenance, protocols payload, RA execution and administrative tasks such as monitoring processes of SMDs/tasks status, heartbeat messages, among others.

Since SMDs resources provision is performed with finite energy sources, it is relevant to faithfully reflect the energy consumption of the resource utilization in order to avoid obtaining from an SMD more or less resources than it is actually able to provide. One methodology to represent resource exhaustion is through in-vivo laboratory tests, where energy consumption happens in the real SMDs that compose the experimental testbeds [44,45,49,81,89]. This methodology requires the setup of a wireless network for allowing SMDs to communicate, the installation of a middleware-like software on SMDs from where the RA logic operates and the execution of real/artificial workloads on SMDs. By nature, such methodology includes baseline energy consumptions, e.g., those derived from connectivity maintenance, SMDs OS and middleware runtime, including the last, code for RA logic, message processing, fail recovery, among others.

An alternative methodology commonly adopted for representing resource exhaustion is through simulation, which can be performed by employing a synthetic or a trace-based scheme. In either scheme, the approaches for quantifying resources availability can be done through tasks or resources characterization information. In the first case, tasks are associated to fixed amounts of energy consumption values, one for each type of resource utilized

Table 10
Implementing aspects of resource exhaustion singularity.

Aspect	Details			RA works
Resource exhaustion accounting methodology	In-vivo laboratory tests			[44,45,49,81,89]
	Simulated	Synthetic driven by	Tasks characterization	[60,61]
			Resource characterization	[54,62,65]
		Trace-based driven by	Tasks characterization	[44,47,49,88]
			Resource characterization	[50,53,55,57]
Consideration of baseline energy consumptions	Yes			[44,45,49,50,55,57,81,88,89]
	No			[47,60–62,65]

during execution. These are, for instance, the amount of energy consumed by I/O operations, the amount of energy consumed by CPU utilization, etc. Since SMDs present heterogeneous hardware performance, these amounts are different for each candidate SMD where the task could execute. For simulating resource exhaustion with this approach, the SMD remaining energy is decremented in the value, i.e., Joules, that the task execution is expected to consumed in the SMD. In synthetic schemes, as in [60,61], the tasks energy values are artificially defined. In trace-based schemes, like [44,47,49,88], energy values are obtained through profiling tasks execution on real SMDs. For example, in [44], the energy consumption of a distributed object recognition application is profiled in seven SMDs including smartphones, tablets and laptops. In [47] a power monitor hardware is used to profile the energy consumption value of tasks characterized by different combinations of data transferring and computing operations in two smartphones. In [88], the average energy consumption value is calculated for a data mining algorithm running over a dataset of 800 Kbytes in a smartphone. In [49], one set of experiments utilize energy values resulting from profiling of a face detection and a speech-to-text applications on two smartphones.

When resource exhaustion is driven by resource characterization, resources utilization units are associated with energy consumption units (e.g., Joules per MIPS, Joules per KB transferred) and tasks are associated to resource utilization units (e.g., amount of float-point operations, amount of input/output data). With this approach, tasks requirements are independent of the SMDs hardware, and for simulating resource exhaustion, SMDs energy is progressively modified considering the *TaskResourceUtilizationUnits* $*$ *SMDEnergyWastedPerResourceUtilizationUnit* equation. In synthetic schemes driven by this approach, e.g., [54,62,65], energy consumption is expressed as a fixed rate per resource utilization unit and those rates are artificially defined. By contrast, with a trace-based scheme [50,55,57], resources energy consumption is derived from real SMDs resource usage.

In [53,55,57], the traces derive from a profiling procedure which involves sampling energy level drops reported by the OS at a fixed CPU usage. The procedure starts with an unplugged fully charged SMD, and ends when the cut-off voltage is reached. Traces for different CPU usages are obtained, e.g., when the device CPU is idle, lightly or heavily used. The traces are exchanged during a simulation to represent the way battery depletion occurs while the SMD is executing a CPU-intensive task, as well as, to represent energy consumption derived from OS execution and owner interaction. In [50], the resource profiling or, as called by the authors, the *energy model* of an SMD, encompasses the usage of several resources at once. The profile is built from a set of benchmarking tests that exercise different resources usage combinations. Each combination is expressed as a linear equation modeling the energy value obtained through the measurement of the energy consumption performed during the execution of the benchmarks. By solving the linear equation system derived from all tests, the energy consumption rate of each resource under consideration is obtained. Baseline energy consumption derived from OS processes are included in these rates. The simulation software uses such SMDs energy model to reflect tasks energy consumption that are described by a resource demand vector with the quantities of each resource needed.

5.4. Findings summary

We dedicate this section to summarize the findings arisen from the singularities-driven analysis used to organize the surveyed RA efforts. The findings also cover aspects which are not central to the categories proposed such as practical value of RA mechanisms and efforts aligned with incentives for encouraging SMD owners participation.

User mobility as an advantage in particular kinds of ad-hoc networks: In works targeting OppNets and DTNs ad-hoc like networks, where the SMDs are likely to spend more time disconnected from each other than connected, it is observed that for computing workload distribution and results collection, the RA logic uses information of SMDs' movements, stops and future locations. Such decision making process frequently considers SMDs exhaustible resources, which seems reasonable since contacts between SMDs would not occur if SMDs energy is not well managed during the time SMDs are not in contact.

User mobility avoided in RA mechanisms that operate from infrastructure-based networks: RA mechanisms using infrastructure-based networks are designed to operate under the coverage range of the BS or a group of BSs. A common strategy to avoid unfinished workload units, caused by user mobility, is score SMDs reliability/availability using information of connection/disconnection events or movement-related parameters, e.g., node direction and node speed. A strong limitation noticed in many of these approaches is that resource exhaustion parameters, such as remaining battery, are not included in the reliability/availability score.

Conservative resource exhaustion characterizes RA mechanisms targeting ad-hoc networks: The energy conservation, or coupled energy utilization strategy, in networks where nodes reachability is maintained by nodes which operate with batteries is a top priority concern in the design of a RA mechanism because overload a hub node too early, means not only loose its computing resources but the connection with other group of computing providers. In DTNs and OppNets, the intermittent nodes connection caused by user mobility doubles the complexity of achieving such energy conservation. Relegate energy conservation in favor to meet tasks deadlines has also been studied [49].

Decoupled energy utilization characterizes RA mechanisms operating with dedicated communication infrastructure: Giving that SMDs do not have to play the role of data forwarders, when RAs operate via infrastructure-based communication, the energy management strategies typically adopted by RAs is focused on harvesting each SMD energy individually. In other words, the wide view to all computing resource providers accessed with a BS that operates with infinite electricity, is seen as an enabler for making a greedy, or decoupled, energy utilization. An energy management policy like this could be associated to an HPC-like usage of resources. In this context, the consideration of a coupled energy utilization policy, i.e., group-aware energy utilization, has not been studied in the literature. Providing that SMD users are frequently connected to infrastructure-based networks, would such a group-aware energy utilization be an incentive for encourage users to cooperate?

Poor attention to lack of ownership related concerns: A scarce quantity of works incorporate lack of ownership-related concerns into the RA logic design or consider it within the evaluation variables (see efforts addressing LO dimension in Table 4). In other words, most of the proposed RA solutions are not prepared, or at least, the performance has not been measured considering the heterogeneity derived from non-dedicated resources. It suggests that during the time resource exploitation occurs, e.g., while the SMD is connected to an AP, up to reach the energy quota offered by the owner, or during the time the completion of assigned tasks last, external tasks are likely to execute with exclusive privileges on the usage of SMDs resources. Such exclusiveness might bring unwanted effects in SMD owner experience like low responsiveness or delay of routines which run in background, e.g., application updates. In relation to this, RA mechanisms addressing a balance between external tasks performance and SMD owner's experience has not been proposed. Achieving such balance represent a relevant combination that does make sense when resource scavenging time is likely to overlaps with the time the owner interacts with his/her SMD. A situation like this is expected to occur when owner willingness to contribute with resources is extended for long time periods, e.g., hours, days, weeks or even undefined. Possible reasons for which this combination and, in general, lack of ownership singularity has receive little attention is due to the difficulty to engage volunteers to participate from experiments, the absence of specific models to simulate owners' behavior in tools commonly used in the area and the lack of SMDs owners characterization in different resource exploitation contexts.

Variability of RA evaluation methodologies: A small proportion of works [44,45,49,81,89] evaluate RA performance with in-vivo tests, i.e., mounting distributed environments using real SMDs. Most of the performance evaluation is involves the configuration of testbeds in simulated environments. With in-vivo tests, results obtained are close-to-reality. However, in-vivo tests suffer from limited experimental setup, and from hard to recreate experimental conditions. Simulation has been adopted so far in the research field. However, a simulation tool that provides all necessary abstractions to combine all aspects mentioned in the cited works is not available. Different simulation tools are used, each one with its own learning curve, and designed for attending the needs of related research areas. For instance, works targeting ad-hoc like networks use OMNet++[2] or NS2 simulator [99], which have models to simulate specific features of wireless communication research field. Others extend software traditionally employed in Grid Computing research, e.g., GridSim [100], SimGrid [101] to support nodes mobility or exhaustible resources with synthetic models that oversimplified battery depletion behavior with a lineal function. There are also works, using custom instantiations of generic simulation environments [54,60,87] -e.g., JavaSIM,[3] MATLAB[4] - or ad-hoc simulation solutions [55,59] which, in either case, documentation is scarce and the source code is not publicly available for reuse and modification.

Another aspect concerning the evaluation of RA mechanisms is the inclusion of communication related parameters. There are RA works [58,59,62,64,66] targeting infrastructure-based networks contextualize their proposals and/or performance evaluation indicating communication aspects between SMDs and the RA logic, i.e., communication technology stack or standards used, e.g., 802.11, WiMax, Cellular networks. However, in general, these aspects have been superficially covered. The works that mostly specify transmission bandwidth, communication range and transmission power, are those addressing mobility issues, which affect wireless links quality that in turn affect the completion time or throughput of tasks. By contrast, in works targeting ad-hoc networking support, the specification of the aforementioned properties, along with latency caused by multi-hop data transmission and routing protocols or even custom RA protocols used are more frequent [19,45–50,81]. It is observed irrespective of whether mobility is the main concern addressed by the RA mechanism. The specification of these parameters in these works relates with the fact that resource coordination depends exclusively on wireless links maintained by the same nodes providing computing resources.

Abundance of research studies with theoretical value only: Irrespective of the networking support targeted, singularity/ies combination addressed and methodology used for evaluation, many state-of-the-art RA mechanisms, which aims at exploiting computing capabilities of SMDs, require hard to obtain tasks information to operate, e.g., execution time. This limits the applicability of all those RA mechanisms because, in the general case, to predict task execution time, it is first necessary to know whether the task actually ends, which means, solving the halting problem [102]. Another input typically required by theoretical RA mechanisms is the energy that tasks consume on every candidate node. Precise models to estimate battery consumption are based on complex differential equations [103,104], and solving them represent a computationally complex task itself.

Poor comparison between RA proposals: It is very unfrequent to see performance comparisons among RA proposals of different authors. This fact can be attributed several causes, some of which were presented previously, like the differences in singularities addressed, networking supports targeted, pursued objectives and different evaluation methodologies used. However, a cause not mentioned yet and observed in various works, is the poor detailed description of variables and/or range of values used in simulated testbeds. There are works, published in prestigious forums, that do not even reference the simulation software employed, or omit important details such as the technique used to simulate the effects of resource exhaustion. All this, compromise the reproducibility of experimental scenarios and discourage the intentions of cross-validating/compare the performance of RA proposals.

Incentive mechanisms limited to theoretical studies: From the all the reviewed works it is observed that economic models for compensating SMDs resources usage and encourage owner participation is a topic that deserves more attention by the research community. Only few works, from those analyzed in this survey, address the concern [45,59,60]. In [45], the authors outfit the RA proposed with a virtual credit mechanism to reward SMDs owners who contribute with resources. A set of experiments were designed to show the impact in energy consumption of SMDs with selfish and altruist owners profiles. In [59,60] the price of resources is defined by means of a non-cooperative bargaining game played by resource providers and resource consumers, and intermediated by resource brokers. In all cases, the price/reward for the usage of resources is obtained by assuming that tasks resource usage, i.e., execution time, energy spent, is known before the actual usage occurs which turn the approaches hard to be practiced.

6. Future research directions

The analysis of the literature review presented in this work evidences an active research community focused in studying resource allocation techniques for scavenging the computing potential of groups of SMDs. In spite of the significant progress made towards

2 https://omnetpp.org/.
3 http://javasim.ncl.ac.uk>.
4 https://www.mathworks.com/products/matlab.html.

understanding how unique characteristics of SMDs – here named singularities – influence on the effective exploitation of SMDs resources, there is plenty of room for improvements in several directions that are detailed below.

Integrate SMD owner behavior and context-aware information into resource allocation decisions

When thinking in SMDs collaborating in the execution of compute-intensive software, several questions arise including when/where such collaboration is likely to occur (user context), and how computing resource scavenging will affect the normal SMDs' owners usage experience, punctually those acting as resource providers. Being unaware of such concerns is a characteristic of most state-of-the-art RA proposals, but this limits their applicability to specific user situations where the owner is forced to explicitly indicate his/her willingness to contribute resources. The envisioning of an always present willingness, automatically enabled and disabled, plus low invasiveness requires the design of adaptive resource scavenging mechanisms. In this sense, context-aware information that integrate user positioning data [96,105,106], human battery charge profiles [107], applications interaction patterns [108] and future owner needs, can be exploited to create indicators to be used in RA decisions. This would contribute not only to increase the effectiveness and scope of resource scavenging but to offer a seamless resource provision experience to SMD owners. In line with this, metrics for quantifying owner interaction perturbation and incentive mechanisms are also necessary.

Work on the practicability of RA approaches

To operate, many RA mechanisms need to be fed with estimations of the computing time of tasks on every candidate node. Achieving accuracy on such estimations is hard, when not impossible for the general case [102]. In practice, the difficulty stems not only from the heterogeneity given by the variability of hardware and base software performance that could coexist in a mobile Grid, but also from the fluctuating performance caused by the effect of singularities discussed in this survey. Therefore, the reported effectiveness of many RA proposals has a theoretical rather than a practical value. Besides, only few proposals targeting similar objectives are compared against each other in terms of performance. In consequence, the following future work lines arise:

- Designing RA mechanisms that, despite relying on static schedules and hard to obtain tasks information to operate, incorporate logic for detecting changes in resources availability and perform actions to adapt the original schedule to such changes, or at least evaluate the impact on the execution of the original plan. Actions taken might include guidelines and indicators for evaluating resource scavenging of nodes joining the group and nodes whose resources were under-estimated, or mitigating failure/overestimation of node resources. Support for resource availability changes must consider costs of detection logic and actions plans in terms of computational complexity, memory usage, and data to be transferred.
- Designing practical RA mechanisms able to operate with incomplete tasks information, and deal with unbalanced load caused by changes in the resources availability.
- Promoting the comparison among theoretical RA mechanisms, practical ones, and among theoretical and practical approaches. This would help to stress and hence to get deep insights on the performance of RA mechanisms under different resource exploitation settings.

Unifying and enhancing the evaluation methodologies

In the performance evaluation and/or comparison of RA mechanisms, complex relationships among components of a mobile Grid – nodes dynamics, tasks characteristics, networking concerns – are involved. Reproducibility of experimental scenarios is possible when experimental variables can be controlled. Besides, for obtaining solid results of the effect of a variable on performance, any study should comprise several values of the variable in a discrete or continuous form. Simulation can aid in achieving such evaluation requirements. However, a unified open source simulation project that gives the community the possibility of modeling all dynamics of such a complex system still does not exist. A project like this, with integrated modeling capabilities such as those listed below, would contribute to speed up, reuse and organize the progress in the area, as in Information Retrieval and High-Performance Computing research fields-:

- Trace-based and synthetic modeling of SMDs battery depletion derived from multiple resources usage, i.e., CPU, screen, data transferring operations.
- Trace-based and synthetic modeling of SMDs resource sharing by multiple users, i.e., CPU, GPU, bandwidth, memory and storage utilization by processes/transfers which belong to different users and run with different priorities.
- Trace-based and synthetic modeling of user mobility, i.e., variability in data transferring time due to connection/ disconnection effect, signal strength, hop count, ability to support various routing and delay tolerant message delivery protocols.
- Abstractions to model RA mechanisms that operate in ad-hoc and infrastructure-based networks.
- Abstractions to model credit-based resource accounting mechanisms to rule resource provision.
- Abstractions for simulating different tasks execution models – including independent tasks and DAG –, and tasks requirements – MIPS, size of input/output data, memory and storage –.

In addition to a unified simulation core, the application of data mining algorithms for filtering, formatting and including data from up-to-date mobile device datasets — e.g., Device Analyzer [108], CRAWDAD [93] and Human Activity Recognition using Smartphones [109] – are required to build ready-to-use, close-to-reality simulation models.

Diversity in networking supports

When differentiating RA works according to the networking support used to operate, it was observed a balanced number of RA efforts designed for infrastructure-based and ad-hoc networks. This is reasonable given the widely spectrum of technologies and standards developed for both types of communication infrastructures. Notwithstanding, performance evaluation should consider communication related concerns, since not only tasks completion time and tasks throughput are affected, but also the data transferring of resources information used by RA mechanisms to decide tasks destination.

Moreover, given the necessity of sustainable and green solutions to support the increasing worldwide data production, and faster data transmission rates requirements, more research efforts are needed in RA mechanisms able to operate in hybrid networking contexts such as those expointing 5G communication technologies [110–112]. These technologies re-think the way networking services are provided by combining Device-to-Device and Device-to-Infrastructure data transferring capabilities. The high data rates promised by such hybrid networking technologies would enable new application scenarios, e.g., stream processing harvesting SMDs' computing, transferring and sensing capabilities [86].

Addressing incentive, security and privacy concerns

An orthogonal concern that needs to be addressed to ensure that SMD collaboration for compute-intensive tasks goes mainstream is how to systematically lure users into offering computing resources and how to reward for the resources used. A strong limitation of market-based resource provisioning proposed in the literature is that credits for resource usage are determined before RA takes place and based on information of tasks requirements, which is hard to estimate. Resource usage accounting mechanisms able to operate in absence of such tasks information are needed. Besides, different accounting mechanisms, adapted to different system objectives and tasks execution requirements, are needed. For instance, nodes could earn credits according to the time or energy employed to complete a task. Interactive applications, for instance, could adopt a time-based reward mechanism, while green applications processing delay-tolerant tasks, an energy-aware reward mechanism.

Other orthogonal concerns are the potential security treats, and privacy preservation, which call for mechanisms that assure transparency in the usage of resources and prevent misusing effects for both counterparts, i.e., resource providers and resource consumers. Resource consumers should be assured that tasks results are authentic and input data, will not be altered neither by intruders in a network nor by malicious applications or the particular SMD that offer resources. From the resource providers perspective, tasks execution is expected not to misuse permissions granted by the owner to steal or alter owner's personal information or violate his/her privacy. Data encryption [113] and operating system level virtualization [114] can help in satisfying these requirements, however, the time and energy overhead associated have not been studied in the context of RA mechanisms.

Consider the Function as a Service (FaaS) concept to leverage SMDs resources

To validate simulation results, several prototypes have been implemented [81,115–117]. However, a publicly available robust middleware capable of exploiting devices computing capability for a wide spectrum of real-life settings is necessary. Such middleware should be designed to deal with all challenges posed by the singularities, while including varying communication technologies, support for different tasks types, and featuring services for resource discovery, load balancing and fault tolerance. In addition, considering the limitations and volatility of SMDs resources, the minimization of execution overhead is an aspect that cannot be left aside.

Lightweightness and stateless code are concepts on which the FaaS model builds on [118]. *Functions* are the building blocks of such serverless computing model, and a simplified way to benefit from distributed computing without having to deal with the complexity of cluster management and configuration tools [119]. New research opportunities emerge from the adoption of this execution model for allowing SMDs to offer/request computing capabilities from SMDs in the vicinity. An SMD willing to offer unused CPU cycles would announce a list of compute-intensive functions at the time it enters and connects to a known network or the owner voluntarily joins an ad-hoc network with other SMDs. In this context, functions might inherit the same discoverability and understandability issues of traditional Web Services [120], aspects that are also worth to investigate.

7. Conclusions

Motivated by the sustained growth of SMDs number and capabilities, the integration of such devices as resource providers to high performance computing Grids and edge networks has been in the agenda of several researchers, since at least the last decade. The BOINC project for Android devices[5] is an example of how such idea started to be a reality in the hand of citizens who want to voluntarily contribute CPU cycles to scientific projects. Nowadays, more applications can benefit from such integration. The fertile field of IoT applications, including smart buildings, smart cities, health-care monitoring, just to mention a few, are hungry of in-the-field low-latency computing resources. The development of RA mechanisms to exploit computing capabilities of nearby SMDs can aid in satisfying such needs. To make this possible, many efforts has been focused on studying and proposing RA mechanisms. In this survey, we studied more than twenty of these efforts to deal with the challenges imposed by the unique features of resource provision with SMDs, i.e., singularities according to our terminology.

The issues caused by the singularities which until now have received more attention by the research community are user mobility and resource exhaustion, while the least attended are the issues related to the non-dedicated nature of SMD resources. To some extent, the communication support used to coordinate SMDs resources influences the way singularities affect the RA process. For instance, in ad-hoc DTN or OppNet-like networks the intermittent connection between SMDs caused by user movement is treated as a normal condition under which resource allocation should be performed. User movement favors encounter SMDs to distribute tasks and collect results. This is not the case in infrastructure-based mobile networks, where SMDs are expected to stay within the coverage area of some fixed communication antenna. In this context, user movement increases the chances of loosing connection, i.e., it is associated to the undesired effect of loosing control over the resources being managed. Then, a common strategy of RA mechanisms operating from infrastructure-based networks is to rate SMDs movement for discarding or assigning less tasks to those SMDs having unreliable connectivity.

In relation to SMDs exhaustible resources, it is common in RA works targeting ad-hoc networks to conserve energy of SMDs as long as possible in order to preserve network reachability among nodes. This is because the energy of a single SMD might fuel the results transfer of many tasks computed by other SMDs that are not directly reached by the SMD who waits for the tasks completion. In other words, this energy fuels not only such tasks execution but also the tasks and results distribution. Indeed, distribution is facilitated in infrastructure-based networks because SMDs contributing with resources are connected through one-hop links to the resource allocation component. Since energy depletion of an SMD is expected not to affect other SMDs, at least in absence of workflow-oriented applications that have tasks dependencies, the finite energy is administrated independently for each SMD.

With regard to issues related to the non-dedicated nature of resources, it has been barely addressed in the context of infrastructure-based networks and hence it needs more attention in the future. The awareness of shared resources has been used exclusively for the benefit of external tasks, i.e., without paying attention about the possible impact on the energetic needs of the SMD owner.

Aside from the challenges that distinguish RA mechanisms for different types of networking support and singularities addressed, it is remarkable the lack of practical RA approaches and unified validation platforms for simulating realistic conditions. A high percentage of the reviewed RA mechanisms require hard to obtain

5 https://play.google.com/store/apps/details?id=edu.berkeley.boinc.

information of tasks characteristics, i.e, without such information they are unable to operate. Moreover, simulation – which is a useful and accepted practice for recreating complex scenarios – is quite often utilized in the area but with synthetic parameters that, to make things worse, are not standardized within the community. This fact and the lack of a unified simulation platform discourage the healthy competition between competing approaches that share similar objectives, so this should be addressed to avoid slowing down the development of the area.

Acknowledgments

We acknowledge the financial support by CONICET through grant PIP 2017–2019 code 11220170100490CO and ANPCyT through grant code PICT-2013-0464. The first author acknowledges his Ph.D. scholarship granted by the CONICET.

References

[1] Malte Kühnemund, Qingshan Wei, Evangelia Darai, Yingjie Wang, Iván Hernández-Neuta, Zhao Yang, Derek Tseng, Annika Ahlford, Lucy Mathot, Tobias Sjöblom, et al., Targeted dna sequencing and in situ mutation analysis using mobile phone microscopy, Nature Commun. 8 (2017) 13913.
[2] Tassaneewan Laksanasopin, Tiffany W. Guo, Samiksha Nayak, Archana A. Sridhara, Shi Xie, Owolabi O. Olowookere, Paolo Cadinu, Fanxing Meng, Natalie H. Chee, Jiyoon Kim, et al., A smartphone dongle for diagnosis of infectious diseases at the point of care, Sci. Transl. Med. 7 (273) (2015) 273re1–273re1.
[3] Min Chen, Ping Zhou, Giancarlo Fortino, Emotion communication system, IEEE Access 5 (2017) 326–337.
[4] Adrian R.G. Harwood, Alistair J. Revell, Parallelisation of an interactive lattice-boltzmann method on an android-powered mobile device, Adv. Eng. Softw. 104 (2017) 38–50.
[5] Francisco Javier Prego, Luis Nieto, Mercedes Solla, Iván Puente, A mobile android tool for simplified gpr data processing in construction applications, Autom. Constr. 89 (2018) 170–182.
[6] Arturo Aquino, Ignacio Barrio, Maria-Paz Diago, Borja Millan, Javier Tardaguila, vitisberry: An android-smartphone application to early evaluate the number of grapevine berries by means of image analysis, Comput. Electron. Agric. 148 (2018) 19–28.
[7] Abdullah Yousafzai, Abdullah Gani, Rafidah Md Noor, Anjum Naveed, Raja Wasim Ahmad, Victor Chang, Computational offloading mechanism for native and android runtime based mobile applications, J. Syst. Softw. 121 (2016) 28–39.
[8] Paulo A.L. Rego, Philipp B. Costa, Emanuel F. Coutinho, Lincoln S. Rocha, Fernando A.M. Trinta, Jose N. de Souza, Performing computation offloading on multiple platforms, Comput. Commun. (2016).
[9] Saeid Abolfazli, Zohreh Sanaei, Erfan Ahmed, Abdullah Gani, Rajkumar Buyya, Cloud-based augmentation for mobile devices: motivation, taxonomies, and open challenges, Commun. Surv. Tutor. IEEE 16 (1) (2014) 337–368.
[10] Niroshinie Fernando, Seng W. Loke, Wenny Rahayu, Mobile cloud computing: A survey, Future Gener. Comput. Syst. 29 (1) (2013) 84–106.
[11] Jieyao Liu, Ejaz Ahmed, Muhammad Shiraz, Abdullah Gani, Rajkumar Buyya, Ahsan Qureshi, Application partitioning algorithms in mobile cloud computing: Taxonomy, review and future directions, J. Netw. Comput. Appl. 48 (2015) 99–117.
[12] Ejaz Ahmed, Abdullah Gani, Muhammad Khurram Khan, Rajkumar Buyya, Samee U. Khan, Seamless application execution in mobile cloud computing: Motivation, taxonomy, and open challenges, J. Netw. Comput. Appl. 52 (2015) 154–172.
[13] Asma Enayet, Md Abdur Razzaque, Mohamma Mehedi Hassan, Atif Alamri, Giancarlo Fortino, A mobility-aware optimal resource allocation architecture for big data task execution on mobile cloud in smart cities, IEEE Commun. Mag. 56 (2) (2018) 110–117.
[14] Nikola Rajovic, Paul M. Carpenter, Isaac Gelado, Nikola Puzovic, Alex Ramirez, Mateo Valero, Supercomputing with commodity cpus: Are mobile socs ready for hpc? in: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, ACM, 2013, p. 40.
[15] Juan Manuel Rodriguez, Alejandro Zunino, Marcelo Campo, Introducing mobile devices into grid systems: a survey, Internat. J. Web Grid Serv. 7 (1) (2011) 1–40.
[16] Jochen Furthmüller, Oliver P. Waldhorst, A survey on grid computing on mobile consumer devices, in: Handbook of Research on P2P and Grid Systems for Service-Oriented Computing, IGI-Global, 2010, pp. 313–363.
[17] Colin Funai, Cristiano Tapparello, He Ba, Bora Karaoglu, Wendi Heinzelman, Extending volunteer computing through mobile ad hoc networking, in: Global Communications Conference, GLOBECOM, 2014 IEEE, IEEE, 2014, pp. 32–38.
[18] Michael. Black, William. Edgar, Exploring mobile devices as grid resources: Using an x86 virtual machine to run boinc on an iphone, in: 10th IEEE/ACM International Conference on Grid Computing, IEEE, 2009, pp. 9–16.
[19] Ting Shi, Mei Yang, Xiang Li, Qing Lei, Yingtao Jiang, An energy-efficient scheduling scheme for time-constrained tasks in local mobile clouds, Pervasive Mob. Comput. 27 (2016) 90–105.
[20] Emiliano Miluzzo, Ramón Cáceres, Yih-Farn Chen, Vision: mclouds-computing on clouds of mobile devices, in: Proceedings of the Third ACM Workshop on Mobile Cloud Computing and Services, ACM, 2012, pp. 9–14.
[21] Gonzalo Huerta-Canepa, Dongman Lee, A virtual cloud computing provider for mobile devices, in: Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and beyond, ACM, 2010, p. 6.
[22] Lin Yang, Wenfeng Li, Masoud Ghandehari, Giancarlo Fortino, People-centric cognitive internet of things for the quantitative analysis of environmental exposure, IEEE Internet Things J. (2017). http://dx.doi.org/10.1109/JIOT.2017.2751307. (in press).
[23] Gianluca Aloi, Giuseppe Caliciuri, Giancarlo Fortino, Raffaele Gravina, P. Pace, Wilma Russo, Claudio Savaglio, Enabling iot interoperability through opportunistic smartphone-based mobile gateways, J. Netw. Comput. Appl. 81 (2017) 74–84.
[24] Haifeng Xu, William O. Collinge, Laura A. Schaefer, Amy E. Landis, Melissa M. Bilec, Alex K. Jones, Towards a commodity solution for the internet of things, Comput. Electr. Engrg. 52 (2016) 138–156.
[25] Marco Conti, Chiara Boldrini, Salil S. Kanhere, Enzo Mingozzi, Elena Pagani, Pedro M. Ruiz, Mohamed Younis, From manet to people-centric networking: Milestones and open research challenges, Comput. Commun. 71 (2015) 1–21.
[26] Marcel C. Castro, Andreas J. Kassler, Carla-Fabiana Chiasserini, Claudio Casetti, Ibrahim Korpeoglu, Peer-to-peer overlay in mobile ad-hoc networks, in: HandBook of Peer-to-Peer Networking, Springer, 2010, pp. 1045–1080.
[27] DaeWon Lee, HwaMin Lee, DooSoon Park, Young-Sik Jeong, Proxy based seamless connection management method in mobile cloud computing, Cluster Comput. 16 (4) (2013) 733–744.
[28] Jiangtao Zhang, Hejiao Huang, Xuan Wang, Resource provision algorithms in cloud computing: A survey, J. Netw. Comput. Appl. 64 (2016) 23–42.
[29] Hameed Hussain, Saif Ur Rehman Malik, Abdul Hameed, Samee Ullah Khan, Gage Bickler, Nasro Min-Allah, Muhammad Bilal Qureshi, Limin Zhang, Wang Yongji, Nasir Ghani, et al., A survey on resource allocation in high performance distributed computing systems, Parallel Comput. 39 (11) (2013) 709–736.
[30] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, Erdal Cayirci, Wireless sensor networks: a survey, Comput. Netw. 38 (4) (2002) 393–422.
[31] Klaus Krauter, Rajkumar Buyya, Muthucumaru Maheswaran, A taxonomy and survey of grid resource management systems for distributed computing, Softw. - Pract. Exp. 32 (2) (2002) 135–164.
[32] Karthik Kumar, Jibang Liu, Yung-Hsiang Lu, Bharat Bhargava, A survey of computation offloading for mobile systems, Mobile Netw. Appl. 18 (1) (2013) 129–140.
[33] M. Sharifi, S. Kafaie, O. Kashefi, A survey and taxonomy of cyber foraging of mobile devices, IEEE Commun. Surv. Tutor. 14 (4) (2012) 1232–1243.
[34] Arani Bhattacharya, Pradipta De, A survey of adaptation techniques in computation offloading, J. Netw. Comput. Appl. 78 (2017) 97–115.
[35] Minhaj Ahmad Khan, A survey of computation offloading strategies for performance improvement of applications running on mobile devices, J. Netw. Comput. Appl. 56 (2015) 28–40.
[36] Flavio Bonomi, Rodolfo Milito, Jiang Zhu, Sateesh Addepalli, Fog computing and its role in the internet of things, in: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, ACM, 2012, pp. 13–16.
[37] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, Lanyu Xu, Edge computing: Vision and challenges, IEEE Internet Things J. 3 (5) (2016) 637–646.
[38] Rodrigo Roman, Javier Lopez, Masahiro Mambo, Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges, Future Gener. Comput. Syst. 78 (Part 2) (2018) 680–698.
[39] Pavel Mach, Zdenek Becvar, Mobile edge computing: A survey on architecture and computation offloading, IEEE Commun. Surv. Tutor. 19 (3) (2017) 1628–1656.
[40] Marjan Gusev, A dew computing solution for iot streaming devices, in: 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO, IEEE, 2017, pp. 387–392.
[41] Karolj Skala, Davor Davidovic, Enis Afgan, Ivan Sovic, Zorislav Sojat, Scalable distributed computing hierarchy: Cloud, fog and dew computing, Open J. Cloud Comput. 2 (1) (2015) 16–24.
[42] Rafael Pérez-Torres, César Torres-Huitzil, Hiram Galeana-Zapién, Power management techniques in smartphone-based mobility sensing systems: A survey, Pervasive Mob. Comput. (2016).
[43] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, Wireless sensor network survey, Comput. Netw. 52 (12) (2008) 2292–2330.

[44] Hariharasudhan Viswanathan, Eun Kyung Lee, Ivan Rodero, Dario Pompili, Uncertainty-aware autonomic resource provisioning for mobile cloud computing, IEEE Trans. Parallel Distrib. Syst. 26 (8) (2015) 2363–2372.

[45] Simin Ghasemi-Falavarjani, Mohammadali Nematbakhsh, Ghahfarokhi Behrouz Shahgholi, Context-aware multi-objective resource allocation in mobile cloud, Comput. Electr. Engrg. 44 (2015) 218–240.

[46] SayedChhattan Shah, Energy efficient and robust allocation of interdependent tasks on mobile ad hoc computational grid, Concurr. Comput.: Pract. Exper. 27 (5) (2015) 1226–1254.

[47] Abderrahmen Mtibaa, Afnan Fahim, Khaled A. Harras, Mostafa H. Ammar, Towards resource sharing in mobile device clouds: Power balancing across mobile devices, in: ACM SIGCOMM Computer Communication Review, vol. 43, ACM, 2013, pp. 51–56.

[48] Sayed Chhattan Shah, Sajjad Hussain Chauhdary, Myong-Soon Park, et al., An effective and robust two-phase resource allocation scheme for interdependent tasks in mobile ad hoc computational grids, J. Parallel Distrib. Comput. 72 (12) (2012) 1664–1679.

[49] Cong Shi, Vasileios Lakafosis, Mostafa H. Ammar, Ellen W. Zegura, Serendipity: enabling remote computing among intermittently connected mobile devices, in: Proceedings of the Thirteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2012, pp. 145–154.

[50] Jochen Furthmüller, Oliver P. Waldhorst, Energy-aware resource sharing with mobile devices, Comput. Netw. 56 (7) (2012) 1920–1934.

[51] Derek G. Murray, Eiko Yoneki, Jon Crowcroft, Steven Hand, The case for crowd computing, in: Proceedings of the Second ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile HandHelds, ACM, 2010, pp. 39–44.

[52] Sameer Shivle, H.J. Siegel, Anthony A. Maciejewski, Prasanna Sugavanam, Tarun Banka, Ralph Castain, Kiran Chindam, Steve Dussinger, Prakash Pichumani, Praveen Satyasekaran, William Saylor, David Sendek, J. Sousa, Jayashree Sridharan, José Velazco, Static allocation of resources to communicating subtasks in a heterogeneous ad hoc grid environment, J. Parallel Distrib. Comput. 66 (4) (2006) 600–611. Algorithms for Wireless and Ad-Hoc Networks.

[53] Matías Hirsch, Juan Manuel Rodriguez, Cristian Mateos, Zunino Alejandro, A two-phase energy-aware scheduling approach for cpu-intensive jobs in mobile grids, J. Grid Comput. 15 (1) (2017) 55–80.

[54] Krishan Veer Singh, Zahid Raza, A quantum-inspired binary gravitational search algorithm–based job-scheduling model for mobile computational grid, Concurr. Comput.: Pract. Exper. 29 (12) (2017).

[55] Matías Hirsch, Juan Manuel Rodriguez, Alejandro Zunino, Cristian Mateos, Battery-aware centralized schedulers for cpu-bound jobs in mobile grids, Pervasive Mob. Comput. 29 (2016) 73–94.

[56] Karim Habak, Mostafa Ammar, Khaled A. Harras, Ellen Zegura, Femto clouds: Leveraging mobile devices to provide cloud service at the edge, in: IEEE 8th International Conference on Cloud Computing, CLOUD, IEEE, 2015, pp. 9–16.

[57] Juan Manuel Rodriguez, Cristian Mateos, Alejandro Zunino, Energy-efficient job stealing for cpu-intensive processing in mobile devices, Computing 96 (2) (2014) 87–117.

[58] JongHyuk Lee, SungJin Choi, JoonMin Gil, Taeweon Suh, HeonChang Yu, A scheduling algorithm with dynamic properties in mobile grid, Front. Comput. Sci. 8 (5) (2014) 847–857.

[59] Mahantesh N. Birje, Sunilkumar S. Manvi, Sajal K. Das, Reliable resources brokering scheme in wireless grids based on non-cooperative bargaining game, J. Netw. Comput. Appl. 39 (2014) 266–279.

[60] Li Chunlin, Li Layuan, Exploiting composition of mobile devices for maximizing user qos under energy constraints in mobile grid, Inform. Sci. 279 (2014) 654–670.

[61] Xianglin Wei, Jianhua Fan, Ziyi Lu, Ke Ding, Application scheduling in mobile cloud computing with load balancing, J. Appl. Math. 2013 (2013).

[62] S. Stephen Vaithiya, S. Mary Saira Bhanu, Zone based job scheduling in mobile grid environment, Internat. J. Grid Comput. Appl. 3 (2) (2012).

[63] Sung Ho Jang, Jong Sik Lee, Mobile resource reliability-based job scheduling for mobile grid, KSII Trans. Internet Inform. Syst. 5 (1) (2011).

[64] Preetam Ghosh, Sajal K. Das, Mobility-aware cost-efficient job scheduling for single-class grid jobs in a generic mobile grid architecture, Future Gener. Comput. Syst. 26 (8) (2010) 1356–1367.

[65] E. Ilavarasan, R. Manoharan, High performance and energy efficient task scheduling algorithm for heterogeneous mobile computing system, Internat. J. Comput. Sci. Inform. Technol. 2 (2) (2010) 10–27.

[66] Sang-Min Park, Young-Bae Ko, Jai-Hoon Kim, Disconnected operation service in mobile grid computing, in: Service-Oriented Computing-ICSOC 2003, Springer, 2003, pp. 499–513.

[67] Donato Macone, Guido Oddi, Antonio Pietrabissa, Mq-routing: Mobility-, gps- and energy-aware routing protocol in {MANETs} for disaster relief scenarios, Ad Hoc Netw. 11 (3) (2013) 861–878.

[68] Álvaro Monares, Sergio F. Ochoa, José A. Pino, Valeria Herskovic, Juan Rodriguez-Covili, Andrés Neyem, Mobile computing in urban emergency situations: Improving the support to firefighters in the field, Expert Syst. Appl. 38 (2) (2011) 1255–1267.

[69] Roberto Aldunate, Gregg Larson, Miguel Nussbaum, Sergio Ochoa, Oriel Herrera, Understanding the role of mobile ad hoc networks in non-traditional contexts, Mobile Wirel. Commun. Netw. (2006) 199–215.

[70] D.C. Marinescu, G.M. Marinescu, Yongchang Ji, L. Boloni, H.J. Siegel, Ad hoc grids: communication and computing in a power constrained environment, in: Proceedings of the 2003 IEEE International Performance, Computing, and Communications Conference, 2003, pp. 113–122.

[71] Cristiano Rezende, Azzedine Boukerche, Richard W. Pazzi, Bruno P.S. Rocha, Antonio A.F. Loureiro, The impact of mobility on mobile ad hoc networks through the perspective of complex networks, J. Parallel Distrib. Comput. 71 (9) (2011) 1189–1200.

[72] Min Liu, Zhongcheng Li, Xiaobing Guo, Eryk Dutkiewicz, Performance analysis and optimization of handoff algorithms in heterogeneous wireless networks, IEEE Trans. Mob. Comput. 7 (7) (2008) 846–857.

[73] Danilo Spano, Fabio Ricciato, Opportunistic time-of-arrival localization in fully asynchronous wireless networks, Pervasive Mob. Comput. (2016).

[74] Wazir Zada Khan, Yang Xiang, Mohammed Y. Aalsalem, Quratulain Arshad, Mobile phone sensing systems: A survey, IEEE Commun. Surv. Tutor. 15 (1) (2013) 402–427.

[75] Francesco Restuccia, Sajal K. Das, Jamie Payton, Incentive mechanisms for participatory sensing: Survey and research challenges, ACM Trans. Sensor Netw. 12 (2) (2016) 13.

[76] Lingjie Duan, Takeshi Kubo, Kohei Sugiyama, Jianwei Huang, Teruyuki Hasegawa, Jean Walrand, Motivating smartphone collaboration in data acquisition and distributed computing, IEEE Trans. Mob. Comput. 13 (10) (2014) 2320–2333.

[77] Davide Vega, Roc Meseguer, Felix Freitag, Sergio F. Ochoa, Effort-based incentives for resource sharing in collaborative volunteer applications, in: IEEE 17th International Conference on Computer Supported Cooperative Work in Design, IEEE, 2013, pp. 37–42.

[78] Etienne Urbah, Peter Kacsuk, Zoltan Farkas, Gilles Fedak, Gabor Kecskemeti, Oleg Lodygensky, Attila Marosi, Zoltan Balaton, Gabriel Caillat, Gabor Gombas, et al., Edges: Bridging egee to boinc and xtremweb, J. Grid Comput. 7 (3) (2009) 335–354.

[79] Antonios Litke, Dimitrios Skoutas, Theodora Varvarigou, Mobile grid computing: Changes and challenges of resource management in a mobile grid environment, in: 5th International Conference on Practical Aspects of Knowledge Management, PAKM 2004, 2004.

[80] Mustafa Y. Arslan, Indrajeet. Singh, Shailendra. Singh, Harsha V. Madhyastha, Karthikeyan. Sundaresan, Srikanth V. Krishnamurthy, Computing while charging: building a distributed computing infrastructure using smartphones, in: Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies, ACM, 2012, pp. 193–204.

[81] Seng W. Loke, Keegan Napier, Abdulaziz Alali, Niroshinie Fernando, Wenny Rahayu, Mobile computations with surrounding devices: Proximity sensing and multilayered work stealing, ACM Trans. Embedded Comput. Syst. 14 (2) (2015) 22:1–22:25.

[82] K. Katsaros, G.C. Polyzos, Optimizing operation of a hierarchical campus-wide mobile grid for intermittent wireless connectivity, in: 15th IEEE Workshop on Local Metropolitan Area Networks, June 2007, pp. 111–116.

[83] Mahadev Satyanarayanan, Mobile computing: The next decade, in: Proceedings of the 1st ACM Workshop on Mobile Cloud Computing and Services: Social Networks and beyond, ACM, New York, NY, USA, 2010, pp. 5:1–5:6.

[84] E. Zola, A.J. Kassler, Minimizing the impact of the handover for mobile users in wlan: A study on performance optimization, Comput. Netw. 107 (2) (2016) 292–303.

[85] Stefano Ferretti, Vittorio Ghini, Fabio Panzieri, A survey on handover management in mobility architectures, Comput. Netw. 94 (2016) 390–413.

[86] Rafael Soto, Carolina Bonacic, Mauricio Marin, Verönica Gil-Costa, Simulating streaming software applications running on clusters of processors and smartphone, in: Lin Zhang, Xiao Song, Yunjie Wu (Eds.), Theory, Methodology, Tools and Applications for Modeling and Simulation of Complex Systems, Springer Singapore, 2016, pp. 180–190.

[87] Bo Li, Yijian Pei, Hao Wu, Bin Shen, Heuristics to allocate high-performance cloudlets for computation offloading in mobile ad hoc clouds, J. Supercomput. 71 (8) (2015) 3009–3036.

[88] Carmela Comito, Deborah Falcone, Domenico Talia, Paolo Trunfio, Energy efficient task allocation over mobile networks, in: IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, DASC, IEEE, 2011, pp. 380–387.

[89] Jua Manuel Rodriguez, Alejandro Zunino, Marcelo Campo, Mobile grid seas: Simple energy-aware scheduler, in: 3rd High-Performance Computing Symposium — 9th JAIIO, 2010.

[90] Haluk Topcuoglu, Salim Hariri, Min-you Wu, Performance-effective and low-complexity task scheduling for heterogeneous computing, IEEE Trans. Parallel Distrib. Syst. 13 (3) (2002) 260–274.

[91] E. Ilavarasan, P. Thambidurai, Low complexity performance effective task scheduling algorithm for heterogeneous computing environments, J. Comput. Sci. 3 (2) (2007) 94–103.

[92] Marco Conti, Silvia Giordano, Martin May, Andrea Passarella, From opportunistic networks to opportunistic computing, IEEE Commun. Mag. 48 (9) (2010) 126–139.

[93] David Kotz, Tristan Henderson, Ilya Abyzov, Jihwang Yeo, CRAWDAD dataset dartmouth/campus (v. 2009-09-09), September 2009. Downloaded from http://crawdad.org/dartmouth/campus/20090909.

[94] Magdalena Balazinska, Paul Castro, Characterizing mobility and network usage in a corporate wireless local-area network, in: Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, ACM, 2003, pp. 303–316.

[95] David Kotz, Kobby Essien, Analysis of a campus-wide wireless network, Wirel. Netw. 11 (1–2) (2005) 115–133.

[96] Veljko Pejovic, Mirco Musolesi, Anticipatory mobile computing: A survey of the state of the art and research challenges, ACM Comput. Surv. 47 (3) (2015) 47.

[97] Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, Deborah Estrin, Diversity in smartphone usage, in: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, ACM, 2010, pp. 179–194.

[98] Raj Jain, Dah-Ming Chiu, William R. Hawe, A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer System, vol. 38, Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984.

[99] Teerawat Issariyakul, Ekram Hossain, Introduction to Network Simulator NS2, in: Springer Science & Business Media, 2011.

[100] R. Buyya, M. Murshed, GridSim: A toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing, Concurr. Comput. Practice Exp. 14 (13) (2002) 1175–1220.

[101] H. Casanova, Simgrid: a toolkit for the simulation of application scheduling, in: Proceedings First IEEE/ACM International Symposium on Cluster Computing and the Grid, 2001, pp. 430–437.

[102] Reinhard Wilhelm, Jakob Engblom, Andreas Ermedahl, Niklas Holsti, Stephan Thesing, David Whalley, Guillem Bernat, Christian Ferdinand, Reinhold Heckmann, Tulika Mitra, Frank Mueller, Isabelle Puaut, Peter Puschner, Jan Staschulat, Per Stenström, The worst-case execution-time problem — overview of methods and survey of tools, ACM Trans. Embedded Comput. Syst. 7 (3) (2008) 36:1–36:53.

[103] Wen-Yeau Chang, The state of charge estimating methods for battery: A review, ISRN Appl. Math. (2013).

[104] Y. Hu, S. Yurkovich, Battery cell state-of-charge estimation using linear parameter varying system techniques, J. Power Sources 198 (2012) 338–350.

[105] Daniel Moreno, Sergio F. Ochoa, Understanding the resource positioning methods that support mobile collaboration, in: Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on, IEEE, 2016, pp. 003676–003682.

[106] Prodromos Makris, Dimitrios N. Skoutas, Charalabos Skianis, A survey on context-aware mobile and wireless networking: On networking and computing environments' integration, IEEE Commun. Surv. Tutor. 15 (1) (2013) 362–386.

[107] Ahmad Rahmati, Angela Qian, Lin Zhong, Understanding human-battery interaction on mobile phones, in: Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services, ACM, New York, NY, USA, 2007, pp. 265–272.

[108] Daniel T. Wagner, Andrew Rice, Alastair R. Beresford, Device analyzer: Large-scale mobile data collection, ACM SIGMETRICS Perform. Eval. Rev. 41 (4) (2014) 53–56.

[109] Davide Anguita, Alessandro Ghio, Luca Oneto, Parra Perez Xavier, Jorge Luis Reyes Ortiz, A public domain dataset for human activity recognition using smartphones, in: Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2013, pp. 437–442.

[110] Tuyen X. Tran, Abolfazl Hajisami, Parul Pandey, Dario Pompili, Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges, IEEE Commun. Mag. 55 (4) (2017) 54–61.

[111] Bego Blanco, Jose Oscar Fajardo, Ioannis Giannoulakis, Emmanouil Kafetzakis, Shuping Peng, Jordi Pérez-Romero, Irena Trajkovska, Pouria S. Khodashenas, Leonardo Goratti, Michele Paolino, et al., Technology pillars in the architecture of future 5g mobile networks: Nfv, mec and sdn, Comput. Stand. Interfaces 54 (2017) 216–228.

[112] Pouria Sayyad Khodashenas, Cristina Ruiz, Muhammad Shuaib Siddiqui, August Betzler, Jordi Ferrer Riera, The role of edge computing in future 5g mobile networks: concept and challenges, in: Cloud and Fog Computing in 5G Mobile Networks: Emerging Advances and Applications, vol. 70, 2017, p. 349.

[113] Syam Kumar Pasupuleti, Subramanian Ramalingam, Rajkumar Buyya, An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing, J. Netw. Comput. Appl. 64 (2016) 12–22.

[114] Sascha Wessel, Manuel Huber, Frederic Stumpf, Claudia Eckert, Improving mobile device security with operating system-level virtualization, Comput. Secur. 52 (2015) 207–220.

[115] A. Perez Campos, J.M. Rodriguez, A. Zunino, An empirical evaluation of a simple energy aware scheduler for mobile grids, in: XLIII CLEI - Inpress Page in Press, IEEE, 2017.

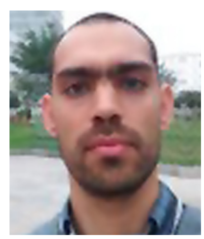
[116] Dimitris Chatzopoulos, Mahdieh Ahmadi, Sokol Kosta, Pan Hui, Openrp: a reputation middleware for opportunistic crowd computing, IEEE Commun. Mag. 54 (7) (2016) 115–121.

[117] Sebastian Schildt, Felix Busching, Enrico Jorns, Lars Wolf, Candis: Heterogenous mobile cloud framework and energy cost-aware scheduling, in: Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing, IEEE, 2013, pp. 1986–1991.

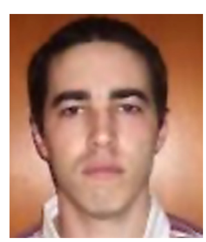
[118] J. Spillner, C. Mateos, D.A. Monge, FaaSter, Better, Cheaper: The Prospect of Serverless Scientific Computing and HPC, in: E. Mocskos, S. Nesmachnow (Eds.), High Performance Computing, in: CARLA 2017. Communications in Computer and Information Science, vol. 796, Springer, Cham, 2018.

[119] Eric Jonas, Shivaram Venkataraman, Ion Stoica, Benjamin Recht, Occupy the cloud: Distributed computing for the 99%, 2017. ArXiv preprint arXiv:1702.04024.

[120] Matías Hirsch, Ana Rodriguez, Juan Manuel Rodriguez, Cristian Mateos, Alejandro Zunino, Spotting and removing wsdl anti-pattern root causes in code-first web services: A thorough evaluation of impact on service discoverability, Comput. Stand. Interfaces (2017).



Matías Hirsch holds a B.Sc. in Systems Engineering and a Ph.D. in Computer Science from UNICEN. He is a teaching assistant at the UNICEN. His research interests include distributed computing using mobile devices farms, modeling and simulation of mobile device computing and mobile/service-oriented computing. Contact him at matias.hirsch@isistan.unicen.edu.ar.



Cristian Mateos has an M.Sc. and a Ph.D. in Computer Science from UNICEN. He is an adjunct professor at the UNICEN and researcher at the CONICET. He is interested in parallel and distributed programming, middlewares, and mobile/service-oriented computing.

Alejandro Zunino has an M.Sc. and a Ph.D. in Computer Science from UNICEN. He is an adjunct professor at the UNICEN and researcher at the CONICET. His research areas include grid computing, service-oriented computing, Semantic Web services, and computer security.