Computer Networks 75 (2014) 297–312

Contents lists available at ScienceDirect

Computer Networks

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

Cooperative proxies: Optimally trading energy and quality of service in mobile devices

Aruna Prem Bianzino, Mikael Asplund, Ekhiotz Jon Vergara, Simin Nadjm-Tehrani *

Department of Computer and Information Science, Linköping University, 581 83 Linköping, Sweden

ARTICLE INFO

Article history:
Received 6 November 2013
Received in revised form 19 June 2014
Accepted 8 October 2014
Available online 17 October 2014

Keywords:
Green networks
Mobile
Energy efficiency
User terminals
Wireless communication energy
Bandwidth sharing

ABSTRACT

This work studies the energy and quality of service (QoS) trade-off in the context of mobile devices with two communication interfaces (a high energy and a low energy interface). We propose an optimisation scheme during underload scenarios where proxy groups are dynamically formed exploiting both interfaces. The scheme integrates a reward mechanism that compensates a proxy while carrying other group members' traffic, and deals with churn (joining and leaving of nodes) in a cell area. For traffic flows that approximate knowledge about current services we show that the scheme can achieve energy savings of 60% for all mobile nodes as whole. We also demonstrate the impact on disruption-sensitive flows as a function of the traffic mix, and that the use of rewards for selection of proxies is a fair mechanism in the long term.

1. Introduction

Networking is intertwined with the fabric of social and professional activities and reliance on mobile communication infrastructures, such as 3rd generation cellular communication (3G), WiFi access points, and future generations of these technologies seems indispensable. Seamless mobility is supported by the cellular infrastructures, which have traditionally been designed with load maximisation and high availability in focus. Ubiquity of smart devices (phones, pads, phablets) has made the above quality of service (QoS) requirements more vivid. The call for sustainable information and communication technology (ICT) has, however, added another dimension to the equation, namely the need for optimising energy consumption.

Several major works focus on reducing the energy footprint of the infrastructure nodes (e.g., [1–3]), which by some estimates constituted 90% of the total communication energy a few years ago [4]. However, with 50 billion devices stipulated to be connected all the time [5], the device energy becomes an important part of the overall energy optimisation. Although device energy use may be small for each unit, with a system perspective, it makes a considerable sum in total.

On top of this contributing factor to the ICT carbon footprint we have an Internet user service availability aspect meaning that the length of uninterrupted use of a device should be maximised for a given user. This essentially amounts to optimising the use of battery resources. From a communication perspective the saving can be achieved in the device within different layers: hardware and firmware (interface card, battery technology), but also software residing in transport and application/session layer (app development, flow management). While vendors will continuously reduce the hardware footprint they cannot optimise energy at packet/flow level since the device is application agnostic. This paper focuses on the *communication* energy optimisation at the *device* level. Other energy saving measures managed within the infrastructure

* Corresponding author.
E-mail addresses: aruna.bianzino@liu.se (A.P. Bianzino), mikael.asplund@liu.se (M. Asplund), ekhiotz.vergara@liu.se (E.J. Vergara), simin.nadjm-tehrani@liu.se (S. Nadjm-Tehrani).

http://dx.doi.org/10.1016/j.comnet.2014.10.013
1389-1286/

(e.g., a base station) are complementary to the approach we propose.

We start by asking the question: under which circumstances can energy be reduced with no major impact on service? The paper starts from the premise that energy savings are possible during intervals of time that a (sub) network is underloaded. In general, usage scenarios consist of periods of underload followed by periods of overload, each demanding their own optimisation regime (one may focus on energy whereas the other typically prioritises throughput). Focusing on periods during which energy savings are possible allows a smart combination of the best of two worlds. A system continues to reap the benefits of accessibility of a high energy access node (AN) within a cell during overloads. However, during underloads, we form multi-user coalitions in that cell using a low energy interface in order to reduce the overall energy in these intervals.[1]

Our approach includes dynamic formation of *proxy groups* whereby one node acts as a proxy which is connected to the high energy interface, and relays the traffic received from/sent to members in the proxy group using a low energy interface. In order to compensate the current proxy for the additional energy used to transmit other nodes' packets, we create a reward-based sharing scheme in which the proxy role rotates in the group. This guarantees some notion of fairness among the nodes in the proxy groups in the long term.

During overloads the bandwidth of an existing proxy group towards the AN will not suffice, requiring reconfiguration of the proxy group formations. This paper addresses the problem of periodic (re)formation of proxy groups such that the overall energy consumed due to data communication is optimised. The local switches within a cell, merging proxy groups when they are underloaded and splitting into more proxy groups when one is overloaded, are automatically dealt with as part of the optimisation algorithm. Note that the degenerate case becomes where all the nodes are their own proxy and carry (only) their own traffic at the high energy interface.

Since variations in load are unpredictable one needs to perform these self-organisations based on the knowledge of the *current* traffic characteristics. Our system treats those flows that are disruption sensitive different from those that have a more relaxed QoS requirement.

The main challenges that need to be addressed are churn (new nodes coming into the cell or leaving the cell), fairness (the node acting as a proxy and thereby spending additional energy is fairly selected over a long time span), and maintained quality of service (coalition switches should preferably not disrupt flows that are sensitive to reconfiguration).

If we compare the classic schemes typically deployed to cope with these issues within networking layers, rotation can be considered as a choice of the forwarding node (at IP/routing layer), while proxy group merge/split can be considered as a means of admission control for maintaining proxy load, or congestion control (at transport layer). Addition of churn to the problem upsets both of these schemes.

First, a fair rotation scheme may not work if the addition of nodes is faster than the rate of feasible rotation. Second, the overhead of frequent reconfiguring (merges and splits) will wipe out the benefits of making coalitions in the first place. Our first attempt at resolving these issues is to characterise a system model under which optimisation can indeed be formulated. This scheme makes attainability of energy savings (at an acceptable level of QoS disruptions) the deciding factor. That is, reconfigurations will only be made if overall energy savings are possible.

The contributions of the paper are as follows:

- An optimisation scheme for dynamically forming proxy groups, that combines a high energy interface used by proxies and low energy interfaces used by proxy group members.
- An integrated rotation scheme, based on accrued rewards and their expenditure, so that the proxy role is shared in a fair manner in a dynamic load scenario.
- A scheme for reconfiguring proxy groups upon load changes that acts as an implicit admission control for the proxy using the high energy interface. Both reconfiguring and rotation (under items 1 and 2 above) take account of disruption sensitivity of flows, thereby minimising disruption to flows that are configuration-sensitive.

In general, the contribution of devices to the energy consumption dynamically changes based on (a) user connections to Internet services – creating non-deterministic flow dynamics, and (b) embedded devices which do not send predetermined packet flows in a time-driven fashion – considered in machine-to-machine or Internet-of-Things scenarios – creating traffic that is event-based and dynamic.

In our traffic model we characterise 4 flow classes covering the current services used by mobile users, and simulate a network that implements the optimisation scheme within a cell. The traffic flows cover applications with soft or hard real-time constraints, or no timing constraints.

Our evaluations (based on 3G and WiFi energy models) show an energy saving of over 60% for all the nodes in the system on average. We then go on to study the usefulness of the reward mechanism as a means of implementing a fairness criteria, and find it to be as intended within a 28 day envelope of operation (i.e., level of unfairness tends to reduce over time). Next we characterise the degree of disturbance to QoS levels offered in a wide range of traffic mixes finding the mixes of traffic for which our scheme makes a valuable trade-off between energy and QoS during underloads.

The structure of the paper is as follows. Section 2 documents the assumptions in our optimisation scheme for the overall energy vs. QoS trade-off. Section 3 describes the

[1] Note that we intentionally use the term *access node* to be technology agnostic. This is a node that is connected to the existing infrastructure and which has a relatively high energy consumption for transmissions, compared to other technologies in the scenario. For example, this is a 3G base station in the model used for evaluations in Section 4, and not a WiFi access point.

optimisation scheme combined with the rotation mechanism. Section 4 is devoted to experimental evaluation of the proposed techniques in a simulated environment and Section 5 discusses potential deployment issues. Related works are presented in Section 6, and Section 7 sums up the conclusions as well as some future directions for work.

2. System model

Our system model is composed of a set N of mobile devices with multiple wireless communication interfaces. These devices are able to use local low energy interfaces to communicate with nearby nodes, as well as the high energy interface to communicate via an access node (AN) which provides Internet connectivity. A proxy group is defined as a group of mobile devices where one of them acts as a proxy for the others. By forming optimal energy-saving proxy groups, the mobile devices are able to reduce their overall communication energy consumption over time.

Our optimisation framework acts on the level of a single AN cell (i.e., where communication is managed by the same AN). The overall scheme can be replicated across an entire infrastructure network, but for the sake of clarity we focus on a single cell in our mathematical model as well as in the evaluation. The optimisation logic and the rotation scheme reside at the AN, which has knowledge of the participating nodes and their current traffic demand. Hence, at a cell level the optimisation is centralised, but the traffic aggregation is dynamic and decentralised. In the remainder of this section we describe the basic assumptions which underpin this work in terms of mobility, proxy capability and traffic model.

2.1. Node mobility

Let $A_t \subseteq N$ denote the dynamic but finite set of mobile nodes that are *active* in the cell at time t. Nodes can enter and leave the cell at any point in time as illustrated in Fig. 1. A node which leaves the cell also has to leave its energy-saving proxy group. This means that the ability of the system to adapt to changes must match the mobility-induced churn (rate of change for A_t).

Let μ represent the average rate at which nodes depart from the cell (i.e., the average time a node spends in the cell is $1/\mu$). We assume that this average departure rate is bounded and not excessively high. Our proposed approach adapts to the changing environment by periodically reconfiguring every T_{step} seconds. This means that the reconfiguration interval should be significantly smaller than the average time a node stays in the system.

$$T_{\text{step}} \ll \frac{1}{\mu} \tag{1}$$

To put this in context, in our evaluations we use 1 min as the value for T_{step} whereas Trestian et al. [6] based on a data set of over 280,000 users report that the average time a node stays with one base station in a 3G context is 40 min.

2.2. Proxy capability

The energy-saving scheme investigated in this paper assumes that every node which participates in the scheme is capable of acting as a proxy for other nearby nodes. Fig. 2 shows the general idea where the proxy nodes route data to and from other nodes in the same cell.

For the purpose of formalisation and evaluation we assume that all nodes within a single cell are able to communicate with each other using the low energy wireless interface. Variations where only some nodes have proxy-capability or shorter-range low-power interfaces can be envisaged. However, we will not use node-to-node communications other than the case where one node in such a pair is a proxy. We are interested in the basic proxy group mechanism and how to optimally form fair and energy-efficient groups. Such principal results are necessary in order to subsequently create derivative schemes to deal with implementation-related problems.

We use the Boolean function $p(i,j) : A_t \times A_t \rightarrow \{0,1\}$ to denote whether j is the proxy for device i (in the figure $p(4,3) = 1$ whereas $p(4,2) = 0$). Every node must be associated with a proxy, and every proxy node must also be its own proxy (e.g., $p(3,3) = 1, p(7,7) = 1$ and $p(10,10) = 1$ in Fig. 2). A trivial proxy is a node that is only its own proxy, e.g., device 1 in Fig. 2 ($p(1,1) = 1$).

Since our focus is on developing a general optimization framework, we only consider high and low energy interfaces in the system model. This high-level abstraction is motivated by the fact that the actual power consumption is technology-dependent (and therefore differs between

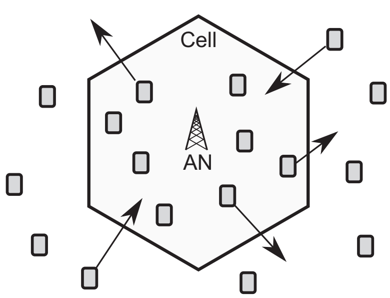

Fig. 1. Node mobility relative to a single cell.

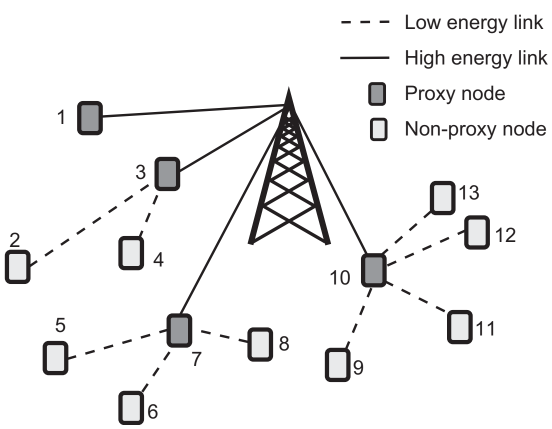

Fig. 2. Example configuration of the system with four proxy groups.

different device models). For the purpose of configuring proxy groups it suffices to assume that a node will consume less energy using the low-energy interface. Thus, the power models are only necessary for evaluation purposes (described in Section 4.1).

Finally, for the proxy scheme to work, the time required to establish a new proxy configuration needs to be significantly smaller than the time between reconfigurations (T_{step}).

2.3. Traffic model

Our proposed approach is concerned with data traffic which we model as bidirectional flows. Each flow is associated with a source node (the mobile device), a data rate, and an expected duration. To reduce model complexity we treat all traffic to and from one device as a single flow (the modelled data rate is the sum of all the uplink and downlink traffic), which means that a node i has at most one flow $l(i)$ representing the *load* for that node in the model. All the generated and received traffic of the mobile devices goes through the AN.

To account for variations on QoS requirements among different flows we consider flows with some real-time constraints as *sensitive* which means that it is undesirable to interrupt these flows even for a short duration (the change of a proxy device is considered as an interruption in a flow). Examples of this kind of flows may be live video streaming and gaming applications. We use the Boolean function $s(i) : A_t \to \{0, 1\}$ to denote whether a node is transmitting or receiving a sensitive flow at time t. The fact that a node is participating in a sensitive flow needs to be known by the AN.

The nodes within a proxy group share the bandwidth B provided by the *high-energy* interface of the proxy node. We assume $l(i) \leqslant B$, considering that a node will not generate more traffic than B when using the low-energy interface. A sudden increase in traffic load will affect the ability of the proxy groups to meet the demand forcing the system to reconfigure. Similarly, a reduction in load will lead to under-utilisation which leads to energy waste. However, when a single node experiences sudden changes in load, by aggregating the load in a proxy group, the impact of these changes is reduced. Given that a per-node load estimate $l(i)$ is maintained for node i during a forecast window T_{for}, the system will be able to adapt to changes in traffic load as long as the reconfiguration interval is not longer than the forecast window: $T_{\text{step}} \leqslant T_{\text{for}}$.

3. Algorithm design

This section presents POEM (Proxy for Optimal Energy saving in Mobile devices), an algorithm which periodically determines a fair, energy and QoS-optimal proxy group configuration. The input to the algorithm is the set of nodes A_t which are active within a cell at time t, the characteristics of the activity they are performing in terms of load $l(i)$ and sensitivity $s(i)$, and the previous proxy group configuration $p_{\text{prev}}(i,j)$. The output is a new optimal and fair proxy configuration $p^*(i,j)$. The algorithm is composed of three steps:

1. Determine an optimal set of proxy groups which minimise flow disruptions and overall energy consumption.
2. Calculate node rewards to reflect how much time each node has spent acting as a proxy for other nodes.
3. Determine a new proxy device for each group using the reward-based fairness scheme.

Finally, since the set of mobile nodes A_t is not static, the system also needs to handle aperiodic events where nodes arrive and depart from the cell between two reconfiguration points. We now proceed to present the three algorithm steps as well as the mechanism to deal with aperiodic events.

3.1. Optimisation

Step 1 is clearly the core part of this algorithm since it determines the optimal proxy group configuration for the next time period. Our optimisation approach is based on two main design principles. The first principle is to prioritise the minimisation of flow disruptions over energy savings. The second principle is the general nature of the algorithm, meaning that it can apply to any kind of devices that present a low and a high-energy interface, without requiring perfect knowledge of the energy characteristics of every single device. As a consequence, the optimal configuration is one with as few high-energy interfaces as possible, i.e., as few and large groups as possible.

In order to perform the proxy group optimisation step, it is necessary to have information about the current traffic flows of the mobile devices. For each node $i \in A_t$ the load estimate $l(i)$ and sensitivity $s(i)$ of the flow originating at node i is an input to optimisation. Recall from Section 2.3, that we assume the flow estimation to be approximately valid for a forecast window of T_{for} which should be greater than the time to the next reconfiguration T_{step}. Note that the system is able to operate also with unreliable predictions (in the evaluation in Section 4 we do not assume the load to stay constant between reconfiguration steps).

The purpose of the optimisation step is to find the set of proxy groups which first of all minimises the amount of proxy switches for nodes with sensitive flows and as a secondary objective minimises the global energy consumption. The outcome of the optimisation step is an optimal but tentative proxy configuration $p_{\text{opt}}(i,j)$. The final choice of proxy within a proxy group can potentially be changed in the subsequent steps of the algorithm to account for the fairness requirement. However, this proxy rotation does not affect the optimality with regards to energy consumption and non-disruption of sensitive flows. The same number of proxies are maintained and the same sensitive flows are respected.

We formulate the optimisation problem as an Integer Linear Programming (ILP) problem with the following objective function:

$$\text{Obj} : \min \sum_{i \in A_t} \left[s(i)\chi(i) + \frac{1}{M_t} p(i,i) \right] \tag{2}$$

where $\chi(i) : A_t \to \{0, 1\}$ is a function taking the value of 1 if the proxy of device i will change with respect to the

Table 1
Optimisation variables and parameters.

Attribute	Domain	Co-domain	Description
$l(i)$	A_t	$\mathbb{R}_0$	Load estimate of device $i \in A_t$ (input parameter)
$s(i)$	A_t	$\{0,1\}$	1 if the flow at node $i \in A_t$ is sensitive to change of proxy (input parameter)
$p_{\text{prev}}(i,j)$	$A_t \times A_t$	$\{0,1\}$	1 if device $j \in A_t$ was the proxy of device $i \in A_t$ in the previous configuration (input parameter)
M_t	n.a.	$\mathbb{N}$	Constant to prioritise sensitive flows, $M_t > \lvert A_t\rvert$ (input parameter)
B	n.a.	$\mathbb{R}_0$	Bandwidth of the high-energy interface (input parameter)
$\chi(i)$	A_t	$\{0,1\}$	1 if the proxy will change for device $i \in A_t$ with respect to the current optimal configuration (optimisation variable)
$p(i,j)$	$A_t \times A_t$	$\{0,1\}$	1 if device $j \in A_t$ is the proxy of device $i \in A_t$ in the current configuration (optimisation variable)

previous proxy configuration (i.e., this term minimises the sensitive flow disruptions).

The second term of Eq. (2) minimises the number of proxy groups (same as the number of devices which are proxy of themselves, $p(i,i)$), corresponding to a minimum system energy consumption. This term is weighted with a factor $1/M_t$ to guarantee that the first term is always prioritised and reflect the strict preference towards solutions with fewer proxy switches for devices with sensitive flows, over solutions resulting in lower energy consumption but disrupting flows. The maximum value that the second term can assume is $|A_t|$ for the degenerate case, while the smallest variation of the first factor is 1 unit. In order to strict prioritise flow conservation over energy saving, M_t should hence be bigger than $|A_t|$. Since any choice of $M_t > |A_t|$ is equivalent here, but a higher value of M_t slows down the exploration of the solution space, $M_t = |A_t| + 1$ represents the best possible choice.

The objective function is subject to the following constraints: (see Table 1)

$$\sum_{j \in A_t} p(i,j) = 1 \quad \forall i \in A_t \tag{3}$$

$$\sum_{j \in A_t} p(j,i) l(j) \leqslant B \quad \forall i \in A_t \tag{4}$$

$$\sum_{i \in A_t} p(i,j) \leqslant |A_t| \cdot p(j,j) \quad \forall j \in A_t \tag{5}$$

$$\sum_{j \in A_t} |p(i,j) - p_{\text{prev}}(i,j)| = 2\chi(i) \quad \forall i \in A_t \tag{6}$$

where Eq. (3) ensures that each device is associated to exactly one proxy. Eq. (4) checks that the total traffic of each proxy group remains within the bandwidth capacity of the high-energy interface of the proxy (i.e., B). Eq. (5) ensures that if a node is acting as proxy for other nodes, then it is acting as proxy also for itself. This is due to the fact that the maximum value that the left side can assume is $|A_t|$. Finally, Eq. (6) guarantees that the variable $\chi(i)$ assumes the correct value, i.e., $\chi(i) = 1$ if device i will change its proxy node, $\chi(i) = 0$ otherwise.

By solving this optimisation problem we obtain a QoS-energy-optimal proxy group configuration $p_{\text{opt}}(i,j)$. However, the optimal solution is not necessarily unique. Within each proxy group, there are several nodes which could potentially act as proxy. We now continue to describe the final steps of the algorithm which is concerned with a fair proxy selection scheme.

3.2. Device reward

The nodes which act as proxies suffer temporarily from an increased energy consumption due to relaying other nodes' data on their high-energy interface. To alleviate this problem we use a reward mechanism to ensure long-term fairness in the cooperation. This mechanism keeps track of the time each device spends acting as proxy for other nodes. Each device i is associated to a reward value $r(i)$ which is a cumulative variable (with an abstract value domain on which increment and decrement are defined), updated in the following way:

- For each time unit spent by a device as proxy of other nodes, its reward value is increased by one unit;
- For each time unit spent by a device using another node as proxy, its reward value is decreased by $1/(G-1)$, where G is the size of the proxy group which the node belongs to.

Within each proxy group this reward policy always sums to zero, so the global reward will tend to stay constant over time. Consider a proxy group of size G, which by definition has 1 proxy node and $(G-1)$ non-proxy nodes. The sum of the reward in the group will be $1 - (G-1) \cdot 1/(G-1) = 1 - 1 = 0$. This holds true also for singular proxy groups where devices act as proxy for themselves and for no other nodes (i.e., not participating in a larger proxy group results in zero reward).

This reward scheme is strictly based on how long time nodes spend as proxies and non-proxies. Other factors such as bandwidth usage or energy cost can be envisaged. However, the energy-cost is not proportional to bandwidth usage and requires complex technology-dependent models to be accurately represented [7].

3.3. Proxy rotation scheme

The defined reward function allows the design of a fair rotation scheme for the role of proxy among devices in a proxy group. The designed rotation scheme starts from the proxy configuration obtained through the optimisation step leading to $p_{\text{opt}}(i,j)$ and works as follows: (i) Every time a reconfiguration is considered, for each proxy group, a check is performed to guarantee that no proxy rotation is applied if some devices in the group have sensitive flows. (ii) If there are no devices with sensitive flows, the device with the lowest value of $r(i)$ inside the group is selected as proxy. The choice is driven by the higher priority we set for not affecting sensitive activities, over reducing the energy consumption. At the same time, the designed reward scheme guarantees a fair rotation in the long term.

When all proxy groups have been considered for a possible non-disruptive proxy switch, the resulting fair and optimal proxy configuration $p^*(i,j)$ can be applied by the mobile devices in the system.[2]

Note that our current model is agnostic to the current energy level in a device. In an implementation it would be possible to adopt a scheme whereby the nodes that are connected to a power source are prioritised when choosing a proxy (though not interesting in evaluations here since our focus is on saving battery charge in mobile scenarios). It would also be possible to use the reading of the current charge level as an input to the selection of the proxy. However, it is worth pointing out that (a) to evaluate that scheme our evaluation here would have to include a generic discharge model for all devices (which is not easy to establish), and (b) the highest-charge node prioritisation would be similar to works that use energy harvesting (in essence opposing the current fairness we achieve in our scheme).

3.4. Aperiodic events

Apart from the periodic reconfiguration, special care is needed to handle the cases of devices arriving or leaving the cell. In particular, when a device arrives in the cell, it forms its own group (for which it is the proxy), and will be normally considered in the system optimisation to join another proxy group at the end of the current T_{step} time window. When a device is leaving the cell, two cases are possible: (i) the device is not the proxy of its proxy group, or (ii) the device is the proxy of its proxy group. In the first case, the device can leave the proxy group without problems and the rest of the configuration is kept until the end of the current T_{step} time window. In the second case, the other devices in the proxy group remain in the same proxy group and a new proxy is selected among them following the rules listed in Section 3.3 (i.e., the device with the lowest reward $r(i)$ will be selected as proxy). Note that in case (ii) some sensitive flows may inevitably be disrupted. However, this will not be due to optimisation-induced changes in the group but rather due to a node physically leaving a cell (the effect of which can be viewed as similar to cellular handover). We will return to this issue in Section 4.2.

4. Evaluation

We evaluate the POEM algorithm in terms of energy savings, degree of QoS assurance and fairness. This section describes the evaluation methodology and settings, and presents the results.

4.1. Evaluation methodology

In order to evaluate the algorithm performance under the specified system model, we define realistic mobility, traffic and energy consumption models. These are plugged into our modular simulation environment. The models are described below.

4.1.1. Mobility model

The proposed algorithm has been evaluated considering a system composed of a single AN cell and a population of devices entering and leaving the system. We consider devices entering the cell following a Poisson process with arrival rate $\lambda = 25 \times 10^{-3}$ Hz, and leaving the system after an exponentially distributed time period with mean $1/\mu = 900$ s. The selected mean period that a device stays in the cell is lower than the mean values reported in the literature (40 min in a 3G context [6]), which is chosen to further stress the system in terms of disruptions.

4.1.2. Traffic model

Modelling traffic flows in terms of duration and bandwidth utilisation in a realistic manner is a difficult task as historical traffic models and characterisations quickly become outdated with the fast evolution of devices. For example, the rapid increase of mobile applications is replacing some of the traffic previously associated with "browsing", thereby changing the traffic characteristics.

We define five distinct groupings of user/application activities on mobile devices leading to five traffic classes. The first classification groups activities leading to traffic flows which are *delay-sensitive*, like VoIP, live radio, etc. The second grouping models the *off-screen* communication, including background traffic, notifications, cloud synchronisation, software updates, etc. This class of flows has no time-constrained characteristics. The third activity type captures *interactive delay-tolerant* traffic sources, like web browsing, instant messaging, applications like Facebook or weather forecast, delay-tolerant games, etc. This class has some application-dependent latency requirements but the nature of the requirements is such that a (short enough) disruption in the flow does not significantly reduce the service quality. The fourth class covers the *delay-constrained* communications. This class describes activities which can tolerate delay up to a point, due to a given buffer policy (e.g., Youtube or Spotify) or to specific timing limitations (e.g., purchasing activities or some games) but the applications are by nature delay-sensitive beyond the toleration point. Finally, the fifth activity type models the periods of inactivity of the device denoted by the *silent* class.

We model each activity in terms of duration, data rate, and the proportion of time a node will spend with this activity on average. The details are reported in Table 2, where "Prop." indicates the average proportion of time spent by each device performing the different activities, "Duration" indicates the minimum and maximum duration of the activity, "Data rate" indicates the minimum and maximum data rates, and "Flow sens." indicates whether the traffic flow is sensitive to changes of the proxy (1 denoting sensitivity). For each activity a data rate value is selected at random as uniformly distributed between minimum and maximum of the indicated interval. Also a duration value is selected at random as uniformly

[2] Note that $p^*(i,j)$ is an optimal solution equivalent to $p_{opt}(i,j)$, as it fulfils the same set of constraints and has the same values for $s(i)\chi(i)$ and the same number of proxy groups, i.e., nodes i for which $p(i,i) = 1$.

Table 2
User application activities and resulting flows.

Activity description	Prop. (%)	Duration (s)	Data rate (kbps)	Flow sens.
Delay-sensitive	5	30–600	30–100	1
Off-screen communication	25	3–20	1–500	0
Delay-tolerant interactive	20	60–3600	50–2000	0
Delay-constrained interactive	10	60–3600	500–2000	1
Silent	40	1–900	–	–

distributed between minimum and maximum of the indicated interval.

Data rate and duration values of the different activities have been selected in order to be representative of all the possible categories included in the groups. As an example, the interactive delay-tolerant group includes activities such as instant messaging or web browsing, which can be characterised by durations as short as one minute, up to durations as long as one hour. At the same time, these activities can result in bandwidth request as low as 50 kbps (e.g., consulting a web page low in media contents) or as high as 2 Mbps (e.g., consulting a web page including videos). Note that bandwidth requests are average values over the activity duration. Instantaneous values can hence include lower minima and higher maxima.

The proportion of time spent by devices performing the different activities has been selected in order to represent the typical behaviour of a user in a non-office, non-home environment. Such behaviour is usually dominated by non-interactive activities (i.e., silent and off-screen), while we consider a mix of the interactive activities mainly including delay-tolerant activities (e.g., delay-tolerant games or Facebook) and a smaller portion of delay-sensitive activities (e.g., live radio or VoIP).

4.1.3. Energy model

Even though the optimization framework is general and does not rely on technology-dependent energy models, we still need to compute the energy savings of our approach. Two widespread wireless interfaces are selected for the evaluation: 3G as the high-energy interface, and WiFi as the low-energy interface. We use a basic energy model in the evaluation to compute the communication energy based on the data rate of the flows resulting from the user activities. A flow represents both the transmitted and the received data.

The communication energy depends on various aspects, such as network inactivity timers, power saving modes or digital modulation schemes [8]. Power consumption per bit decreases when the data rate increases [9], and the energy consumption is not proportional to the amount of data sent [10]. The current state-of-the-art energy models profile the transmission energy at packet level [11–13], which does not satisfy our modelling requirements which requires a flow-based energy model for the evaluation.

The flow-based model is a simplification which estimates the resulting energy consumption for a given traffic flow (composed by duration and average data rate). The model uses parameters based on measurement data [12] together with earlier packet level modelling work [7]. The average power consumption for a flow is modelled as a linear function of its data rate (which for a node i is the same as the traffic load $l(i)$). Eqs. (7) and (8) show the relationship between load and average power consumption for the high and low interfaces (3G and WiFi) respectively. The energy consumption of the flow is calculated by multiplying the average power by the flow duration.

$$P_i^H = 800 + 20 \cdot l(i)\ [\mathrm{mW}] \tag{7}$$

$$P_i^L = 100 + 90 \cdot l(i)\ [\mathrm{mW}] \tag{8}$$

Since the proxy is receiving data from all the other devices in the proxy group, we also account for the energy cost of its WiFi interface during receiving intervals. This model reflects the energy cost of a typical WiFi software access point running in a mobile device [14].

Finally, it is important to note that the absolute values used in the model are not decisive since the energy savings depend on the proportional difference between the high and low energy interfaces. However, earlier physical energy measurements in an implementation of a proxy group in real devices [15], shows that the results are in the same orders of magnitude.

4.1.4. Simulation environment

Our cooperative proxy scheme and system model have been implemented in a custom event-based simulator written in the Python language. This simulator schedules and treats discrete events occurring in the system (e.g., device arrival/departure). Events are scheduled over continuous time. The algorithm execution event happens every T_{step}, while all other events are asynchronous (according to the mobility and traffic models). The simulator has been designed in a modular way so that other optimisation engines can be plugged in, as well as other traffic flow, mobility and energy models.

The implementation of the POEM algorithm uses the Gurobi optimisation engine[3] to find the optimal proxy group configurations. The simulations have been executed on a MacBook Pro retina (2.7 GHz Intel Core i7 processor, 16 GB of DDR3 RAM and OS X 10.8.3).

4.2. Simulation results

We compare the proposed solution against a non-cooperative scenario where each device is independently connected to the AN using its high-energy interface. The solution was tested over an 18 h period. The simulation parameters are summarised in Table 3. The time unit used for calculating the reward is seconds.

Fig. 3 (Top) is reporting the variation over time of the total traffic generated by the active devices in the cell. Horizontal black dashed lines represent the capacity of a high energy interface, so that the first such line over the traffic curve represents a lower bound on the number of proxy groups needed at a given time instant. Fig. 3 (Bottom) reports the variation over time of the number of devices

[3] A Gurobi Python interface is available and documented in [16].

Table 3
Simulation parameters.

Parameter	Symbol	Value	Description
Simulation time	T	18 h	Total simulated time
Reconfiguration period	T_{step}	60 s	Determines how often the algorithm is run
Device population	$\|N\|$	200	Total number of distinct devices considered
Arrival rate	λ	25 mHz	Arrival rate of devices in the cell
Average time in the cell	$1/\mu$	900 s	Average time spent by devices in the cell
Interface bandwidth	B	2 Mbps	Bandwidth of the high energy interface of the devices

in the considered cell, with an average between 20 and 25 devices and peaks of more than 35 devices.

4.2.1. Energy saving

Fig. 4 (Left) reports the variation over time of the total power consumption in one run, accounting for all the active devices in the cell, both for the non-cooperative case and for the POEM case. As we can see, cooperative proxy grouping considerably reduces the overall power consumption of devices. To better quantify the average energy saving, Fig. 4 (Right) reports the average normalised total energy consumption for 10 different runs of the simulation, together with the corresponding standard deviation. The 10 runs are obtained by varying the random seed, i.e., we change the traffic and mobility pattern for the different devices, while keeping the statistical values for traffic profiles constant. The POEM algorithm leads to an energy saving of over 60% of the total energy consumed by the devices.

4.2.2. QoS considerations

Saving energy with no concern for QoS is easy. An infinite delay of a service trivially saves a lot of energy. This section presents our balanced approach to QoS treatment. As an indicator of the QoS for the proposed solution we look at the number of times in which a device's proxy is changed while it is performing a sensitive activity. This can happen in two cases: (i) the proxy leaves the cell ("Proxy Mobility"), or (ii) the total traffic of the proxy group increases and the proxy is no longer able to carry it with the previous configuration, thus a group split has to take place ("Reconfiguration"). In the latter case, the algorithm naturally prefers new configurations in which devices performing non-sensitive activities are relocated to another proxy group, but this is not always possible. The left bar of Fig. 5 reports the total number devices facing a proxy change while performing a sensitive activity, considering all devices and over the entire simulation period in the run depicted in Fig. 5 (Left).

In order to better understand how this metric compares to other mobility-related events we contrast it to the occurrences of devices leaving the cell while performing a sensitive activity ("DeviceMobility", right bar of Fig. 5). This phenomenon is not affected by the algorithm and happens due to the mobility of devices themselves. We thus observe that the proxy changes for sensitive activities introduced by the algorithm are mainly due to device mobility, and similar in number to the occurrences of devices leaving the cell while performing sensitive activities (typically requiring a handover from one AN to another).

4.2.3. Fairness considerations

As an indicator for the fairness of the rotation mechanism in the proposed algorithm, we look at the distribution of the reward among the devices in the device population. Fig. 6 reports the evolution in time for the analysed distribution, starting from a uniform distribution – "Initial" in Fig. 6. The x-axis reports the bottom reward value of each one-hour bin of the distribution, while the y-axis reports the percentage of devices in each bin. As we can see, the algorithm tends to reduce the variation of reward values by forcing the devices with low reward to act as proxy and avoiding devices with high reward to act as proxy. A measure of the change of the reward distribution is given by the standard deviation of the devices' reward, reported as "std" in the graphs in Fig. 6. The standard deviation steadily decreases over time indicating that the reward values progressively concentrate around the average value.

4.2.4. Traffic mix variations

The presence of sensitive flows in the system limits the ability to make arbitrary changes to the proxy group configurations. Therefore, it is interesting to evaluate how the trade-off between energy savings and QoS is affected by the proportion of sensitive flows. For this reason, we evaluate the algorithm with different mixes of the traffic types described in Table 2. In particular, we vary the sum of the sensitive activities (i.e., "Delay-sensitive" and "Delay-constrained") from 0% to 100%, while maintaining their internal proportion, as well as maintaining proportion among the non-sensitive activities.

The corresponding variation in terms of energy saving and proxy changes for sensitive flows is reported in Fig. 7. We can observe that the energy saving accomplished by POEM is slightly decreasing for increasing amounts of sensitive flows until they form around 50% of the total traffic. For higher amounts of sensitive flows in the traffic mix, the energy saving decreases faster down to no savings for 100% of sensitive flows. This behaviour is due to the decreasing freedom in grouping devices with increasing constraints due to sensitive flows.

The number of proxy changes (shown as bars in Fig. 7) for sensitive activities is increasing for growing amounts of sensitive activities, as can be expected. The increase levels off when there is around 80% sensitive flows in the traffic mix due to the fact that only small or singular proxy groups are formed under these circumstances.

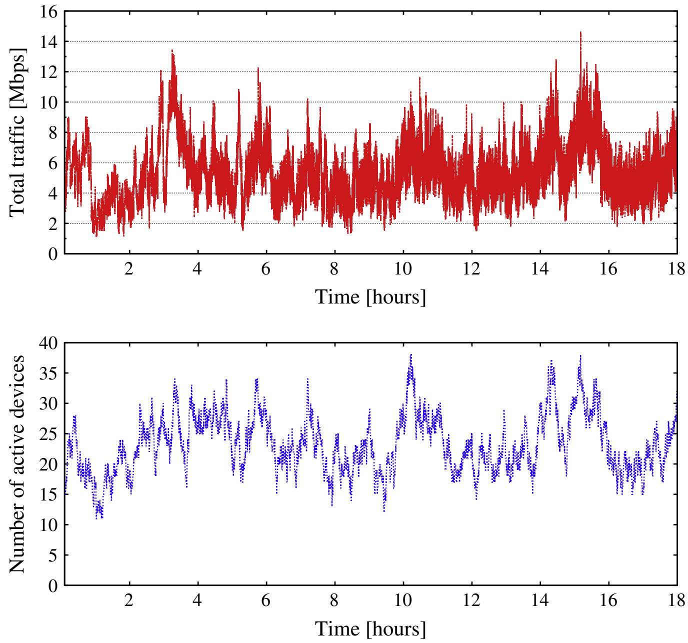

Fig. 3. Total cell traffic (Top) and number of active devices in the cell (Bottom).

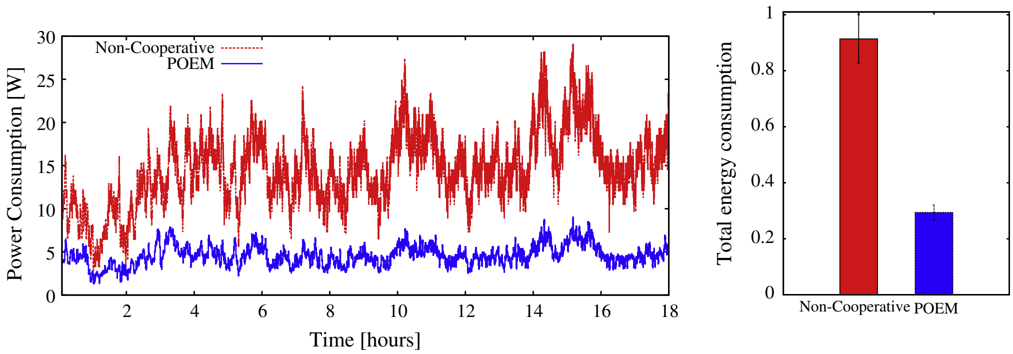

Fig. 4. Total power consumption in one run (Left) and average normalised energy consumption over 10 runs (Right).

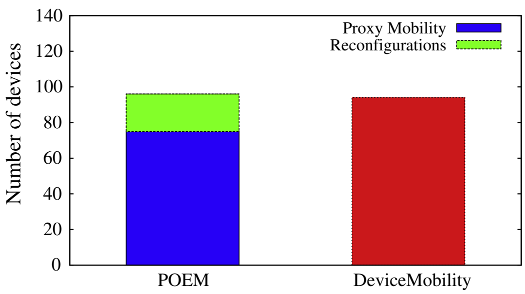

Fig. 5. Proxy switch for devices performing sensitive activities, due to proxy mobility and due to system reconfigurations versus number of devices leaving the cell while performing sensitive activities.

We can hence conclude that the energy saving achieved by our scheme is just slightly reduced by the QoS constraints which prevent sensitive flows from suffering unwarranted proxy changes, and this holds true as long as sensitive flows do not represent the majority of the traffic. At the same time, our scheme can achieve considerable energy savings even when the flows are mostly sensitive.

4.2.5. Prioritisation factor

To further verify that the relative cost of upholding QoS does not decrease the total energy savings, we consider a variation of the original optimisation problem. Clearly, the case where the energy is prioritised over disrupting sensitive flows is not interesting if the energy savings are comparable, and in general both users and network operators will strictly prioritise QoS over energy saving.

Nevertheless, for completeness of the analysis, the priority between factors in the optimisation function may be tuned by introducing a parameter α (with values between 0 and 1) in Eq. (2). The objective function will thus change as follows:

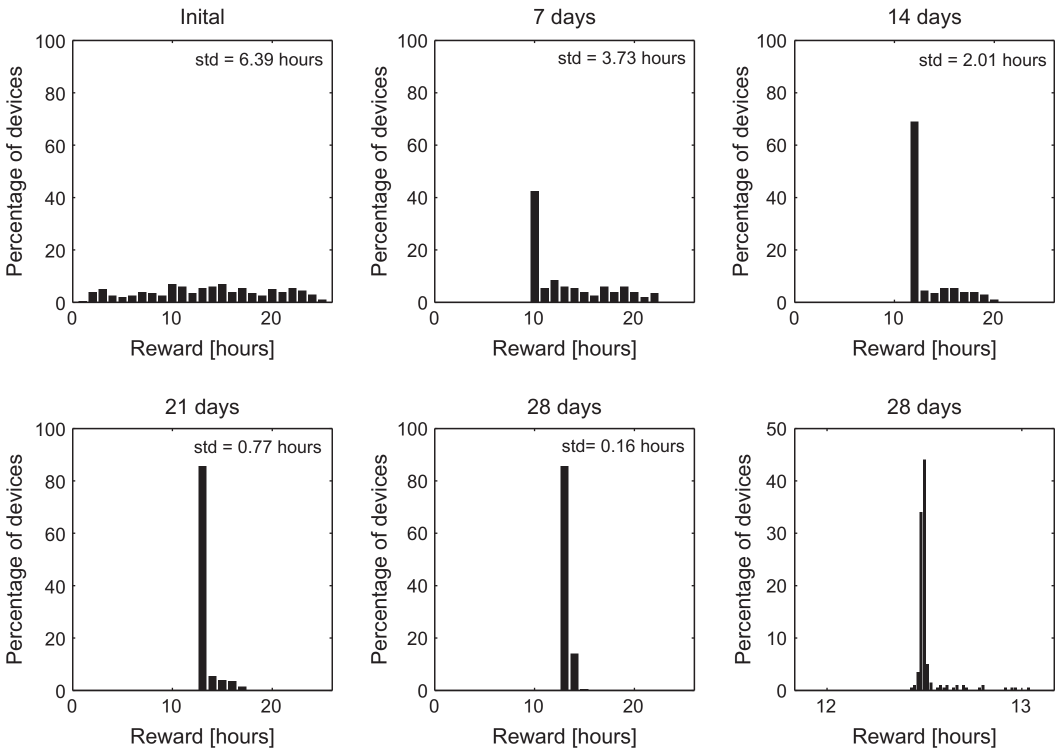

Fig. 6. Evolution of the reward distribution, respectively: initial distribution (top left), distribution after one week of simulation (top centre), distribution after two weeks of simulation (top right), distribution after three weeks of simulation (bottom left), distribution after four weeks of simulation (bottom centre). A detailed view of the distribution after 28 days is also reported (bottom right).

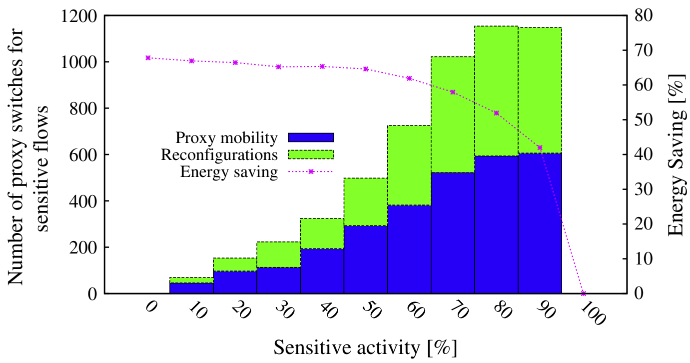

Fig. 7. Variation of the proxy switch for sensitive flows and of the energy saving for increasing amount of sensitive activities.

$$\text{Obj}: \ \min \sum_{i \in A_t} \left[s(i)\chi(i) + \frac{1}{M_t^{\alpha}} p(i,i) \right] \tag{9}$$

While keeping M_t constant at $M_t = |A_t| + 1$, when $\alpha = 1$, Eq. (9) is identical to the original objective function (Eq. (2)), i.e., minimizing the number of disrupted sensitive flows is prioritised over the energy consumption. When $\alpha = 0$ no prioritisation is performed. We compute the average number of disruptions and energy savings for three different values of α. The results based on 300 optimizations for each α value are reported in Fig. 8.

This shows that the average energy savings stay nearly constant while the disruptions are increased compared to the original case ($\alpha = 1$). Thus, the optimisation with lower values of α is inferior to the original POEM approach where QoS is prioritised over energy savings.

5. Potentials for deployment

This section discusses some practical implementation aspects of our proposed scheme.

Our evaluation set up exploited energy models from 3G cellular and WiFi. Other possibilities exist and the basic

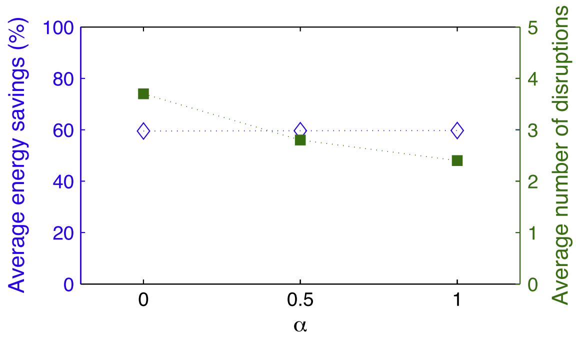

Fig. 8. Average energy savings and sensitive flow disruption for different α values.

optimisation is orthogonal to those choices. Example of potential technologies for enabling the proxy group are WiFi-Direct [17], Bluetooth Personal Area Networks (PANs) [18] or LTE device-to-device communication [19]. The proxy group management overhead depends on the selected technology (i.e., signalling bandwidth and energy, reconfiguration time). An implementation based on Android phones has provided a preliminary evaluation showing that a simple rotation (i.e., just the proxy switch operation) in a proxy group of 2 phones uses 1.56 J of energy due to signalling in the device which is going to be proxy, and 0.8 J of energy due to signalling in the other device. Thus, there are indications that the approach would be worthwhile pursuing in practical evaluations [10].

How to elaborate the cooperation between the devices and the AN in a practical setting is a technological aspect. There exist technologies that allow traffic routing through a relay node, such as software access points already available in mobile devices [20]. Seamless management of the proxy role change requires an extension to the current technology. A possible solution may consist of a virtual access point in each device which is remotely activated and configured by the AN with the optimal operational settings.

The long term accounting and management of rewards is also an aspect that needs attention in a future deployment. The reward level of the different devices needs to be maintained by a service provider. Per device information is already stored in the home location register database (HLR) in cellular networks. This can be extended to include the reward information.

In cellular networks of today, per node traffic measurements (which are used when allocating radio resources) are already available at the base station by the Radio Link Control protocol [21]. Thus, we do not envisage major issues with tracking current classification of sensitive/not-sensitive traffic.

Another issue worthy of attention in an optimisation setting is the convergence time. The time to run the algorithm and compute the optimal configuration should be much smaller than the time for which the configuration is kept (i.e., T_{step}). In our general purpose simulation environment, the optimisation execution time has been on average 162.24 ms, with a standard deviation of 59.15 ms. This is much smaller than the $T_{step} = 60$ s we consider.[4]

Finally, an important issue to consider is the time taken for each reconfiguration. The reconfiguration time should be much smaller than the time for which the configuration is kept (i.e., T_{step}). A practical implementation of a bandwidth sharing mechanism has been described in [10], which is using off-the-shelf technologies and results in a reconfiguration time of about 4 s (with no effort towards making the implementation optimised at all). This is one order of magnitude smaller than the $T_{step} = 60$ s we consider.

6. Related works

In recent years energy awareness has become an substantial concern in networking, notably driven by economical, environmental and marketing factors. Considerable amount of effort has been directed towards the reduction of unnecessary energy consumption in fixed networks. Bianzino et al. [22] survey the main approaches in wired networks such as on–off techniques for routing devices and ports, adaptive-link rate or energy-aware routing. Similarly, optical networks have attracted a great deal of attention [23,24].

6.1. Energy-efficient cellular networks

Mobile network operators are naturally interested in saving energy, mostly due to the increasing operational cost of their networks [25,26]. New proposals often adopt infrastructure-centric approaches rather than considering the energy consumption of user terminals [1–4,27]. A comprehensive description of the challenges on the infrastructure side of green cellular networks is provided on a survey by Hasan et al. [28]. Several works employ optimisation methods (e.g., integer linear programming or meta-heuristics) to reduce the energy consumption of the infrastructure in different contexts (e.g., wireless mesh networks [29], cellular network planning [30], WLANs [31] or cellular networks [32,33]). Most approaches combine on–off schemes for infrastructure nodes (e.g., access points) considering capacity demand and user association. The work by Mahapatra et al. [34] proposes a general framework to manage green communications in heterogeneous networks with different radio access technologies. These works could act in a complementary manner to our work where association of a proxy to an access node (as opposed to another) would be subjected to additional optimisation on the infrastructure side. We believe that handset energy is also worthwhile to optimise and that works like ours are needed as a first step towards a holistic approach to green communication.

The focus of our work is on reducing the communication energy consumption of multiple nodes leveraging a second lower power interface. The readers interested in other energy aspects than communication for single

[4] The average optimisation execution time has been computed over the 1080 optimisations executed in the 18 h of simulated time.

Table 4
Summary of the works considering cooperative techniques for mobile devices.

Ref.	Context	Approach	Technology	Evaluation method	Metric	Results
[48]	Clustering	Game theory (iterative Prisoner's Dilemma)	GPRS/WiFi	Simulation (Netlogo)	Energy savings	56%
[43]	Web download	Divide the list of Web items to download among nodes	Cellular/ Bluetooth	Testbed (Nokia N70)	Reduced download time	47%
[45]	Content distribution	Distribute the content via P2P communication in the cluster	LTE/device-to-device (D2D)	Simulation	Energy savings	0–59%
[46]	Content distribution (streaming)	Cooperative cellular P2P streaming via clustering	3G/WiFi ad hoc and 3G/Bluetooth	Simulation (3G/WiFi ad hoc), Testbed (3G/ Bluetooth)	Extended system lifetime	8 times (simulated), 27% (measured)
[47]	Content distribution	Opportunistic cellular data offload via target-set selection	Cellular/short-range (WiFi or Bluetooth)	Simulation and prototype	Energy is not the focus	–
[18]	Clustering	Protocol for WiFi creating Bluetooth clusters based on heuristics	WiFi/Bluetooth	Simulation and testbed	Energy savings	48–61%
[10]	Clustering	Preliminary energy studies for a node clustering architecture	3G/WiFi	Packet-level simulation (real traffic traces and operator settings)	Energy savings	12–68%
[49]	Cellular overloads	Schedulers for uplink packet forwarding during overloads	LTE/WiFi	Numerical	Energy savings	Up to a factor of 2.5
[50]	Hotspot creation	An hotspot architecture to reduce data via a cloud proxy and mobile cooperative download	3G/WiFi	Testbed	Energy savings	38–71%
[15]	Clustering	Physical measurements	3G/short-range (WiFi or Bluetooth)	Testbed (LG Nexus 4 E960)	Energy savings	55–63%

mobile devices are referred to the survey by Vallina–Rodriguez and Crowcroft [35].

6.2. Wireless ad hoc networks

While forming ad hoc configurations as opposed to using an existing infrastructure is a widely studied topic [36,37], energy consumption in idle listening mode is so far the Achilles' heel of ad hoc communication [38,39]. Thus, most cooperative techniques employ clustering methods resulting in one-hop communication in a star topology.

6.3. Energy-efficient clustering

Wireless Sensor Networks (WSN) is the area in which the energy efficiency within clustering has been deeply studied. Compared to the WSN literature [40–42], mobile nodes in a cellular network lack a common goal (in contrast to data collection in WSN). This necessitates mechanisms to ensure fairness and incentives to cooperate such as our reward-scheme. Moreover, the footprint characteristics of the different interfaces and the traffic load for the individual nodes is different and unpredictable.

Cooperative techniques considering mobile devices often employ a mix of wireless interfaces with different energy footprint characteristics, and reduce the aggregated energy consumption of multiple nodes. Table 4 summarises main aspects of the works categorised as cooperative techniques, which are further described below.

Several works employ clustering assuming that the nodes are interested in the same content. Perrucci et al. [43] present a cooperative mobile web browsing approach combining short-range and cellular networks. The mobile phones cooperatively use their unutilised cellular links to increase the downlink data rate using the user thinking times. The work focuses on reducing download time, claiming that it reduces the total energy consumption as well. Al-Kanj et al. [44] extend the previous work by analytically studying the impact of network parameters on the energy consumption and group formation of nodes interested in the same content. Yaacoub et al. [45] group mobile devices in clusters using device-to-device (D2D) communications to share content of common interest. Chen et al. [46] propose a cooperative video streaming system which reduces the number of cellular links and redundant transmissions by sharing the downloaded data via short-range interfaces. The work by Han et al. [47] proposes to employ long and short-range interfaces to opportunistically perform information delivery with the goal of cellular traffic offloading. The information is first delivered to a target set of users using the long-range interface, which is opportunistically propagated to the rest via the short-range interface. We believe our work is the first instance of cooperative schemes for energy saving and with unrelated user traffic or goals.

Yoo et al. [18] present a cooperative approach to reduce the energy consumption of WiFi creating Bluetooth clusters. The clusters are created based on heuristics given the available bandwidth per node. Our work has a similar aspiration, but minimises the energy consumption considering also the impact on sensitive flows. Perrucci et al. [48] introduce a MAC layer scheme to improve the energy efficiency of cooperation in wireless communication. Earlier work by two of the present authors [10] experimentally studies the energy savings in node coalitions using real user traces. The current work generalises the scheme to organise the proxy groups and minimise the total energy cost as well as incorporating the QoS considerations. A short exposure to experimental evaluation of cluster head scheduling (uplink packet forwarding) within a bi-radio (WiFi–LTE) scenario is provided by Asadi et al. [49], which assumes fully utilised links. This preliminary study indicates that an optimising scheme like ours with a reward-based sharing scheme is worthwhile to pursue in a LTE setting.

Cool-Tether [50] is a WiFi hotspot that offloads the energy burden of the WiFi AP from mobile devices to a laptop, and uses a cloud-based server to reduce the amount of data sent. Our work minimises the total reduction of energy reserve of mobile nodes, including QoS and churn considerations.

Some works [51,52] focus on improving the performance in terms of bandwidth, allowing the nodes experiencing good link quality relay the traffic from the poorly connected ones. This results in saving energy by shortening the transmission times and improving the media access.

Although not strictly related to our optimisation framework, the following groups of works are related to the use of several interfaces or relaying traffic e.g., by software access points:

6.4. Link selection algorithms

Several works consider the problem of selecting the most efficient interface at a certain point in time based on energy [53–55] or QoS criteria [56–58]. These works are orthogonal to our scheme since the link selection algorithms run in each user device, and do not consider cooperation among devices.

6.5. Energy-efficient software access points

Recent mobile devices are able to set up on-the-fly software access points (SoftAP) leading to clustering. Since the AP coordinates the communication, the clients can employ power saving mechanisms. The techniques to reduce the cost of the AP are typically similar to the infrastructure-centric techniques. DozyAP [14] focuses on reducing the energy consumption of WiFi tethering and shows that the WiFi interface of a SoftAP can sleep up to 88% of the time. Camps-Mur et al. [17] study and propose power management protocols to reduce the energy consumption of SoftAP using WiFi Direct. Keshav et al. [59] focus their efforts on coordinating the power saving techniques provided by the cellular wireless communication to the ones provided by WiFi to extend the battery life of the SoftAP. These techniques complement our work by reducing the energy cost of the proxy in our scheme.

6.6. Cognitive radio

Radwan et al. [60] present the main scheme of the C2POWER project based on the intersection of cooperation, heterogeneous networks and advanced short range communication, using the context information provided by cognitive radios. Our work does not consider cognitive radio, but the interested reader is directed to the available surveys, e.g., [61,62].

6.7. Security and privacy

While the energy saving potential of cooperative techniques is promising, security and privacy issues are commonly the main impediments that hold back the applicability of the techniques. So far, we have addressed the fairness and the incentives for engaging in a cooperative scheme. In this context, Lei et al. [63] analyse the business models in the case of operator controlled peer-to-peer communication in combination with LTE.

To sum up, compared to the other works our work proposes a novel optimal algorithm that greatly reduces the energy consumption of user terminals while considering the impact of sensitive flows.

7. Conclusions

Techniques for energy efficiency range over those that minimise the energy footprint of computations, and those that minimise the impact of communication. With respect to the latter, earlier approaches have focused on how one device can maximise the use of the available channels by elaborate channel allocation schemes and efficient utilisation of bandwidth. Paradoxically, more efficient channel utilisations together with the advent of smart phones also encourages massive data transmissions with high energy consumption. Our work has distinguished the need to (a) focus on energy saving potentials during intervals that systems are underloaded, and (b) act cooperatively so that the reduced energy saving can be used for the benefit of all. Both of these are in line with the current philosophies of participatory action.

The novelty of our work is to show that a scheme whereby multiple radio interfaces are optimally exploited in a proxy group creates both energy efficiency and accountability of the share of each device by allocation of rewards and expenditure of them. A major aspect of our work has been to include the QoS requirements as first class citizens in the optimisation criteria and to illustrate that mobility (churn) and avoidance of flow disruptions can be accommodated in the scheme. Our conclusion is that flows with soft or hard real-time requirements do not need to be disrupted with this cooperative scheme, any more than they would in current scenarios where churn essentially disrupts the flows if not managed by a cellular handover. We have also shown that fairness can be achieved in sharing the individual devices' battery resources for the benefit of all, leading to a 60% energy saving.

The proposed algorithm can be implemented in operator networks and is compatible with the signalling data collection with respect flow types and charging structures. Furthermore, an extension of the described mechanism to work across multiple cells is straightforward and allows full operability across a provider network.

7.1. Future works

Extensions of the work would need to evaluate the energy footprint of the scheme more thoroughly, although our preliminary studies look promising [10,15]. Enriching the energy model by adding the impact of mobility on signal strength would be a direct extension. Extending the current results to a network level with different cells is also interesting.

Other directions of work would try to limit the level of disruptions further by special schemes that differentiate the nature of current flows in more detail, e.g., by considering not only the lost flow as a single "call drop" but in terms of its accumulated utility over time (along the lines of earlier work [64]).

New variations of the reward scheme would be interesting to study, including voluntarily acting as a proxy to gain rewards, or considering other handset factors in determining the next proxy. Comparing with fairness-oblivious schemes that resemble energy harvesting mechanisms, whereby a node with a higher current charge in the battery (or indeed connected to a power source) would be preferred when choosing a proxy, is an interesting exercise to pursue. Our current reward scheme which is time-based could be extended with more elaborate schemes taking into account bandwidth usage and contribution to the energy cost of the proxy. It remains a challenge to design such a scheme which is both sufficiently general and which would be perceived by users as fair.

Furthermore, fully distributed schemes are worthy of detailed studies, both in terms of the comparative overhead with the current AN-based scheme, and for combining with more elaborate reward schemes, e.g., those based on social networking.

Further studies can also theoretically consider the limits of the energy-QoS trade-off, i.e., the conditions under which the dynamic reconfiguration would not manage to save (enough) energy at an acceptable cost to QoS.

Acknowledgments

This work has been supported by the Swedish national Graduate school in computer science (CUGS). The authors would like to thank the anonymous reviewers for their insightful comments that helped us to improve the paper.

References

[1] I. Humar, X. Ge, L. Xiang, M. Jo, M. Chen, J. Zhang, Rethinking energy efficiency models of cellular networks with embodied energy, IEEE Netw. 25 (2011) 40–49.

[2] J. Zhou, M. Li, L. Liu, X. She, L. Chen, Energy source aware target cell selection and coverage optimization for power saving in cellular networks, in: Proceedings of the 2010 IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social Computing, GREENCOM-CPSCOM, 2010, pp. 1–8. http://dx.doi.org/10.1109/GreenCom-CPSCom.2010.56.

[3] Z. Niu, Y. Wu, J. Gong, Z. Yang, Cell zooming for cost-efficient green cellular networks, IEEE Commun. Mag. 48 (2010) 74–79.

[4] M.A. Marsan, M. Meo, Energy efficient wireless internet access with cooperative cellular networks, in: Computer Networks, Elsevier, vol. 55, 2011, pp. 386–398.

[5] Ericsson, More than 50 Billion Connected Devices (white paper), 2011.

[6] I. Trestian, S. Ranjan, A. Kuzmanovic, A. Nucci, Measuring serendipity: connecting people, locations and interests in a mobile 3g network, in: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, IMC, 2009, pp. 267–279. http://dx.doi.org/10.1145/1644893.1644926.

[7] E.J. Vergara, S. Nadjm-Tehrani, M. Prihodko, Energybox: Disclosing the wireless transmission energy cost for mobile devices, in: Sustainable Computing: Informatics and Systems, Elsevier, vol. 4, 2014.

[8] N. Ding, D. Wagner, X. Chen, A. Pathak, Y.C. Hu, A. Rice, Characterizing and modeling the impact of wireless signal strength on smartphone battery drain, SIGMETRICS Perform. Eval. Rev. 41 (2013) 29–40.

[9] L. Wang, J. Manner, Energy consumption analysis of WLAN, 2G and 3G interfaces, in: Proceedings of the 2010 IEEE/ACM International Conference on Green Computing and Communications & International Conference on Cyber, Physical and Social Computing, GREENCOM-CPSCOM, 2010, pp. 300–307. http://dx.doi.org/10.1109/GreenCom-CPSCom.2010.81.

[10] E.J. Vergara, S. Nadjm-Tehrani, Watts2share: energy-aware traffic consolidation, in: IEEE International Conference on Green Computing and Communications, GreenCom, 2013. http://dx.doi.org/10.1109/GreenCom-iThings-CPSCom.2013.29.

[11] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, O. Spatscheck, in: Profiling Resource Usage for mobile Applications: A Cross-Layer Approach, MobiSys, ACM, 2011, pp. 321–334, http://dx.doi.org/10.1145/1999995.2000026.

[12] J. Huang, F. Qian, A. Gerber, Z.M. Mao, S. Sen, O. Spatscheck, A close examination of performance and power characteristics of 4G LTE networks, in: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, MobiSys, ACM, 2012, pp. 225–238, http://dx.doi.org/10.1145/2307636.2307658.

[13] A. Pathak, Y.C. Hu, M. Zhang, Where is the energy spent inside my app?: Fine grained energy accounting on smartphones with Eprof, in: Proceedings of the 7th ACM European Conference on Computer Systems, EuroSys, 2012, pp. 29–42. http://dx.doi.org/10.1145/2168836.2168841.

[14] H. Han, Y. Liu, G. Shen, Y. Zhang, Q. Li, DozyAP: power-efficient Wi-Fi tethering, in: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, MobiSys, ACM, 2012, pp. 421–434, http://dx.doi.org/10.1145/2307636.2307675.

[15] U. Moreno Arocena, Energy Consumption Studies for 3G Traffic Consolidation on Android using WiFi and Bluetooth, Master's Thesis, 2014. <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-103393>.

[16] Gurobi, Gurobi Optimization, 2013. <http://www.gurobi.com> (accessed 28.10.13).

[17] D. Camps-Mur, X. Pérez-Costa, S. Sallent-Ribes, Designing energy efficient access points with Wi-Fi direct, in: Computer Networks, Elsevier, vol. 55, 2011.

[18] J.W. Yoo, K.H. Park, A cooperative clustering protocol for energy saving of mobile devices with wlan and bluetooth interfaces, IEEE Trans. Mob. Comput. 10 (2011) 491–504.

[19] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, Z. Turanyi, Design aspects of network assisted device-to-device communications, IEEE Commun. Mag. 50 (2012) 170–177.

[20] Wi-Fi Alliance, Wi-Fi Direct, 2013. <http://www.wi-fi.org/discover-andlearn/wi-fi-direct> (accessed 28.10.13).

[21] RLC protocol, 3GPP Specification 25.322, 2014. <http://www.3gpp.org/ftp/Specs/html-info/25322.htm> (accessed 20.10.14).

[22] A. Bianzino, C. Chaudet, D. Rossi, J. Rougier, A survey of green networking research, IEEE Commun. Surv. Tutorials 14 (2012) 3–20.

[23] W. Hou, L. Guo, X. Wei, X. Gong, Multi-granularity and robust grooming in power- and port-cost-efficient IP over WDM networks, in: Computer Networks, Elsevier, vol. 56, 2012, pp. 2383–2399.

[24] Y. Zhang, P. Chowdhury, M. Tornatore, B. Mukherjee, Energy efficiency in telecom optical networks, IEEE Commun. Surv. Tutorials 12 (2010) 441–458.

[25] M. Etoh, T. Ohya, Y. Nakayama, Energy consumption issues on mobile network systems, in: Proceedings of the 2008 International Symposium on Applications and the Internet, IEEE, 2008, pp. 365–368, http://dx.doi.org/10.1109/SAINT.2008.84.

[26] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, A. Fehske, How much energy is needed to run a wireless network? IEEE Wireless Commun. 18 (2011) 40–49.

[27] M. Marsan, L. Chiaraviglio, D. Ciullo, M. Meo, Optimal energy savings in cellular access networks, in: International Conference on Communications. ICC Workshops, IEEE, 2009, pp. 1–5, http://dx.doi.org/10.1109/ICCW.2009.5208045.

[28] Z. Hasan, H. Boostanimehr, V. Bhargava, Green cellular networks: a survey, some research issues and challenges, IEEE Commun. Surv. Tutorials 13 (2011) 524–540.

[29] A. Amokrane, R. Langar, R. Boutaba, G. Pujolle, Energy efficient management framework for multihop TDMA-based wireless networks, in: Computer Networks, Elsevier, vol. 62, 2014, pp. 29–42.

[30] G. Classen, A.M.C.A. Koster, A. Schmeink, Robust planning of green wireless networks, in: 5th International Conference on Network Games, Control and Optimization, NetGCooP, IEEE, 2011, pp. 1–5.

[31] J. Lorincz, A. Capone, D. Begušić, Optimized network management for energy savings of wireless access networks, in: Computer Networks, Elsevier, 55, 2011, pp. 514–540.

[32] K. Son, H. Kim, Y. Yi, B. Krishnamachari, Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks, IEEE J. Sel. Areas Commun. 29 (2011) 1525–1536.

[33] F. Alaca, A. Sediq, H. Yanikomeroglu, A genetic algorithm based cell switch-off scheme for energy saving in dense cell deployments, in: Globecom Workshops, IEEE, 2012, pp. 63–68, http://dx.doi.org/10.1109/GLOCOMW.2012.6477545.

[34] R. Mahapatra, A.D. Domenico, R. Gupta, E.C. Strinati, Green framework for future heterogeneous wireless networks, in: Computer Networks, Elsevier, vol. 57, 2013, pp. 1518–1528.

[35] N. Vallina-Rodriguez, J. Crowcroft, Energy management techniques in modern mobile handsets, IEEE Commun. Surv. Tutorials 15 (2013) 179–198.

[36] I. Chlamtac, M. Conti, J.J.-N. Liu, Mobile ad hoc networking: imperatives and challenges, in: Ad Hoc Networks, Elsevier, vol. 1, 2003, pp. 13–64.

[37] P. Basu, N. Khan, T.D.C. Little, A mobility based metric for clustering in mobile ad hoc networks, in: International Conference on Distributed Computing Systems Workshop, IEEE, 2001, pp. 413–418, http://dx.doi.org/10.1109/CDCS.2001.918738.

[38] E.J. Vergara, S. Nadjm-Tehrani, M. Asplund, U. Zurutuza, Resource footprint of a manycast protocol implementation on multiple mobile platforms, in: Proceedings of the 2011 Fifth International Conference on Next Generation Mobile Applications, Services and Technologies, NGMAST, IEEE, 2011, pp. 154–160, http://dx.doi.org/10.1109/NGMAST.2011.36.

[39] S. Trifunovic, B. Distl, D. Schatzmann, F. Legendre, WiFi-Opp: ad-hoc-less opportunistic networking, in: Proceedings of the 6th ACM workshop on Challenged networks, CHANTS, 2011, pp. 37–42. http://dx.doi.org/10.1145/2030652.2030664.

[40] W.R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in: Proceedings of the 33rd Hawaii International Conference on System Sciences, HICSS, IEEE, 2000, http://dx.doi.org/10.1109/HICSS.2000.926982.

[41] G. Anastasi, M. Conti, M. Di Francesco, A. Passarella, Energy conservation in wireless sensor networks: a survey, in: Ad Hoc Networks, Elsevier, vol. 7, 2009, pp. 537–568.

[42] O. Boyinbode, H. Le, A. Mbogho, M. Takizawa, R. Poliah, A survey on clustering algorithms for wireless sensor networks, in: 13th International Conference on Network-Based Information Systems, NBiS, IEEE, 2010, pp. 358–364, http://dx.doi.org/10.1109/NBiS.2010.59.

[43] G.P. Perrucci, F.H.P. Fitzek, Q. Zhang, M.D. Katz, Cooperative mobile web browsing, EURASIP J. Wireless Commun. Netw. 2009 (2009).

[44] L. Al-Kanj, Z. Dawy, Impact of network parameters on the design of energy-aware cooperative content distribution protocols, in: Transactions on Emerging Telecommunications Technologies, Wiley, vol. 24, 2013, pp. 317–330.

[45] E. Yaacoub, H. Ghazzai, M.-S. Alouini, A. Abu-Dayya, Achieving energy efficiency in LTE with joint D2D communications and green networking techniques, in: 9th International Wireless Communications and Mobile Computing Conference, IWCMC, IEEE, 2013, pp. 270–275, http://dx.doi.org/10.1109/IWCMC.2013.6583571.

[46] M.-H. Chen, C.-F. Chou, K.-H. Lee, C.-Y. Chang, On cooperative energy-efficient p2p live streaming system for mobile hotspots, in: IEEE International Conference on Green Computing and Communications, GreenCom, 2013. http://dx.doi.org/10.1109/GreenCom-iThings-CPSCom.2013.53.

[47] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, A. Srinivasan, Mobile data offloading through opportunistic communications and social participation, IEEE Trans. Mob. Comput. 11 (2012) 821–834.

[48] G.P. Perrucci, P. Anggraeni, S. Wardana, F.H.P. Fitzek, M. Katz, Bio-inspired energy-aware protocol design for cooperative wireless networks, Int. J. Autonom. Adapt. Commun. Syst., Indersci. 4 (2011) 164–179.

[49] A. Asadi, V. Mancuso, Energy efficient opportunistic uplink packet forwarding in hybrid wireless networks, in: Proceedings of the 4th International Conference on Future Energy Systems, e-Energy, ACM, 2013, pp. 261–262, http://dx.doi.org/10.1145/2487166.2487197.

[50] A. Sharma, V. Navda, R. Ramjee, V.N. Padmanabhan, E.M. Belding, Cool-tether: energy efficient on-the-fly WiFi hot-spots using mobile phones, in: Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies, CoNEXT, ACM, 2009, pp. 109–120, http://dx.doi.org/10.1145/1658939.1658952.

[51] L. Guo, X. Ding, H. Wang, Q. Li, S. Chen, X. Zhang, Cooperative relay service in a wireless LAN, IEEE J. Sel. Areas Commun. 25 (2007) 355–368.

[52] H. Luo, R. Ramjee, P. Sinha, L.E. Li, S. Lu, UCAN: a unified cellular and ad-hoc network architecture, in: Proceedings of the 9th Annual International Conference on Mobile Computing and Networking, MobiCom, ACM, 2003, pp. 353–367, http://dx.doi.org/10.1145/938985.939021.

[53] T. Pering, Y. Agarwal, R. Gupta, R. Want, Coolspots: reducing the power consumption of wireless mobile devices with multiple radio interfaces, in: Proceedings of the 4th International Conference on Mobile Systems, Applications and Services, MobiSys, ACM, 2006, pp. 220–232, http://dx.doi.org/10.1145/1134680.1134704.

[54] A. Balasubramanian, R. Mahajan, A. Venkataramani, Augmenting mobile 3G using WiFi, in: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, MobiSys, ACM, 2010, pp. 209–222, http://dx.doi.org/10.1145/1814433.1814456.

[55] M.-R. Ra, J. Paek, A.B. Sharma, R. Govindan, M.H. Krieger, M.J. Neely, Energy-delay tradeoffs in smartphone applications, in: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, MobiSys, ACM, 2010, pp. 255–270, http://dx.doi.org/10.1145/1814433.1814459.

[56] M. Nam, N. Choi, Y. Seok, Y. Choi, WISE: energy-efficient interface selection on vertical handoff between 3G networks and WLANs, in: Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, vol. 1, PIMRC, 2004, pp. 692–698. http://dx.doi.org/10.1109/PIMRC.2004.1370959.

[57] B.D. Higgins, A. Reda, T. Alperovich, J. Flinn, T.J. Giuli, B. Noble, D. Watson, Intentional networking: opportunistic exploitation of mobile network diversity, in: Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking, MobiCom, ACM, 2010, pp. 73–84, http://dx.doi.org/10.1145/1859995.1860005.

[58] J. Rodriguez, M. Tsagaropoulos, I. Politis, S. Kotsopoulos, T. Dagiuklas, A middleware architecture supporting seamless and secure multimedia services across an intertechnology radio access network, IEEE Wireless Commun. 16 (2009) 24–31.

[59] K. Keshav, V. Indukuri, P. Venkataram, Energy efficient scheduling in 4G smart phones for mobile hotspot application, in: National Conference on Communications, NCC, IEEE, 2012, pp. 1–5, http://dx.doi.org/10.1109/NCC.2012.6176904.

[60] A. Radwan, J. Rodriguez, I. Gomes, E. Sá, C2POWER approach for power saving in multi-standard wireless devices, in: J. Rodriguez, R. Tafazolli, C. Verikoukis (Eds.), 6th International ICST Conference on Mobile Multimedia Communications, MOBIMEDIA 2010, LNICST, vol. 77, Springer, 2012, pp. 440–451, http://dx.doi.org/10.1007/978-3-642-35155-6_35.

[61] B. Wang, K. Liu, Advances in cognitive radio networks: a survey, IEEE J. Sel. Top. Signal Process. 5 (2011) 5–23.

[62] J. Marinho, E. Monteiro, Cognitive radio: survey on communication protocols, spectrum decision issues, and future research directions, in: Wireless Networks, Kluwer Academic Publishers, vol. 18, 2012, pp. 147–164.

[63] L. Lei, Z. Zhong, C. Lin, X. Shen, Operator controlled device-to-device communications in LTE-advanced networks, IEEE Wireless Commun. 19 (2012) 96–104.

[64] C. Curescu, S. Nadjm-Tehrani, A bidding algorithm for optimized utility-based resource allocation in ad hoc networks, IEEE Trans. Mob. Comput. 7 (2008) 1397–1414.

Aruna Prem Bianzino (M'09) is a postdoc researcher at the Real-Time Systems Laboratory in Linköping University, in Linköping, Sweden. He received his M.Sc. degree from the Politecnico di Torino, Italy, in 2008 and his Ph.D. degree from the Telecom ParisTech Institute, in Paris, France, in 2012. His research interests include green networking, traffic engineering and distributed algorithms.

Mikael Asplund is an Assistant Professor at the Real-time Systems Group in Linköping University, Sweden. From 2011–2012 he worked one year as a Research Fellow in Trinity College, Dublin. He received his M.Sc. degree in Computer Science and Engineering in 2005 and his Ph.D. in Computer Science in 2011 both from Linköping University. His Ph.D. thesis focused on design and analysis of partition-tolerant distributed systems, including development of middleware services for maintaining consistency and information dissemination algorithms for disaster area networks. His current research interests include dependable distributed systems, mobile and vehicular computing and real-time systems.

Ekhiotz Jon Vergara is a Ph.D. student at the Real-time Systems Group in Linöping University, Sweden. He received his B.Sc. and M.Sc. degrees in Telecommunication Engineering from Mondragon University, Spain. Since 2011 he has been working on energy-efficient solutions for wireless communication at the user end. His research interests include wireless networking, distributed systems and green computing.

Simin Nadjm-Tehrani received her B.Sc. degree from Manchester University, UK, and did her postgraduate studies leading to a Ph.D. in Computer Science at Linköping University, Sweden, in 1994. During 2006–2008 she was a full professor at University of Luxembourg, and is currently a Professor in Dependable Distributed Systems at Department of Computer and Information Science, Linköping University, where she has led the Real-time Systems Laboratory since 2000. Her research interests relate to networks and systems with dependability requirements and resource constraints.