Energy efficiency and traffic offloading in wireless mesh networks with delay bounds

Rosario G. Garroppo[1], Bernard Gendron[2], Gianfranco Nencioni[1], Luca Tavanti[1]
[1]Dept. of Information Engineering, University of Pisa, Italy
[2]CIRRELT and DIRO, Université de Montréal, Montréal, Canada

Abstract

In this paper we study a wireless access network based on the IEEE 802.11 standard and enriched with features such as caching and mesh networking. This system is analysed in terms of energy efficiency and traffic offloading, two objectives that are somewhat in contrast, but both relevant to network and service providers as they directly impact the operational cost. In addition, quality of service is also accounted for, in the form of guaranteed bandwidth and bounded delay. To this aim, we developed a mathematical model of the system and solved it to optimality by means of integer linear programming. We can thus show how much can be saved both in terms of energy and traffic, also considering various tradeoff points among the two contrasting objectives. As a last step, we provide an investigation on the benefits of adding traffic aggregation features to the mathematical model.

Index Terms

Energy efficiency; traffic offloading; wireless mesh networks; content delivery; caching; quality of service; optimization.

I. Introduction

Medium and large Wireless Local Area Networks (WLANs) are now a popular technique for providing Internet access to the increasing number of mobile user devices. Along with them, many vendors and networks operators have been selling and deploying WLANs enriched with meshing capabilities, also known as wireless mesh networks (WMNs). Recently, there has been an increasing interest in extending the capabilities of both WLANs and WMNs to meet the recent advances in various technological and social fields.

In particular, the steep rise in the fruition and sharing of multimedia contents over the Internet has pushed for the definition and deployment of novel network architectures in order to efficiently deliver such contents to the end users. Content Distribution Networks (CDNs) are currently the dominant way for achieving this goal. By caching and replicating the contents on servers placed near to the network boundaries, CDNs allow to reduce the congestion of the Internet and improve the quality of the delivery service [1], [2].

In order to merge the two worlds – WLANs/WMNs on one side, and CDNs on the other – some researchers have proposed to broaden the CDN model to include also the access point or gateway (GW). The idea is that the GW can be exploited, either as a replacement or as an addition to the CDN, to store the contents destined to the users, thus reducing the congestion of the Internet and improve the quality of the delivery service [2], [3], [4].

Another topical field is energy efficiency, a.k.a. "green" networking. While the majority of the researchers have explored various aspects of cellular networks [5], [6], it is also possible to find some approaches for reducing the power consumption of carrier-grade WLANs [7], [8], as well as for improving the energy performance of the WLAN protocols and procedures [9]. Similarly, some studies have been presented on power saving solutions in wireless mesh networks [10], even though the majority of such studies focus on battery-powered nodes rather than on plugged equipment (see e.g. [11], [12]). Remarkably, there have already been also attemps at merging the CDN and mesh paradigms towards the common goal of energy efficiency, as exposed for example in [13], [14].

Given these premises, we build our work on the system architecture that arises from joining the WLAN access system with the CDN and the mesh paradigms. Specifically, we address a Content-Delivery Wireless Mesh Network (CDWMN), i.e. a wireless LAN in which the GWs, besides granting Internet access to the user terminals (UTs), are also able to perform (limited) content caching and to handle wireless multi-hop paths. In delivering the various multimedia contents to the end-users, the CDWMN should strive to achieve the best energy efficiency and traffic offloading performance. Indeed, both goals are important for the network/service provider, because they both translate in cost saving. At the same time, a minimum level of quality of service (QoS) should also be guaranteed in order to satisfy the customers' expectations.

The main contribution of the paper is therefore an investigation on the potential of the CDWMN from both an energy-aware and a traffic-aware perspective. In detail, we devise a resource allocation and routing scheme (named QETAC, i.e. QoS-based Energy and Traffic Allocation for CDWMN) that jointly considers all the distinctive CDWMN features mentioned above for delivering the contents in an efficient and QoS-compliant manner. The objective is either to minimise the overall energy consumption or to maximise the traffic that can be offloaded from the core network. Since the two objectives are partially contrasting, we study them both singularly and in combination. Accordingly, we formulate a mathematical programming model to analyse the performance of this scheme, especially with reference to the basic (and currently most employed) approach of having a set of independent cache-less GWs. The satisfation of a minimal set of QoS requirements, namely bandwidth and delay, is accounted for throughout the model.

The final goal and result of our work is to quantify the maximum overall power and traffic savings that can be achieved by the proposed QETAC strategy. Yet, in order to make the optimisation practical, both goals are to be achieved under some QoS constraint. In our case, this constraint is the average per-flow delay. We can thus provide an indication on how much a CDWMN can be green and how much traffic it can offload under the optimal configuration. To the best of our knowledge, no previous work has performed a similar study.

II. Related work

The capability and potential of the CDWMN or similar architectures, especially with regard to the energetic issue, has recently received some attention from the research community.

From a general perspective, Mor *et al.* [15] presented a broad classification of energy efficient techniques for WMN, summarising the main features of the considered classes, including design/planning techniques, load balancing approaches, traffic consolidation, energy-aware traffic handover, gateway selection, topology control and energy-aware routing. Jiang *et al.* [16] analyse a content delivery architecture based on geographically dispersed groups of "last-mile" CDN servers, e.g. set-top boxes located within users' homes. The analysis was focused on the design of scalable and adaptive mechanisms to jointly manage content replication and request routing within the presented architecture. With respect to the CDWMN architecture, such last-mile CDN servers are neither shared among the users nor connected by a wireless mesh.

Rossi *et al.* [17] devised some empirical procedures and a distributed protocol for relocating the user terminals of a residential community network[1] (RCN) to the various GWs in order to switch off some GWs and thus save energy. Similarly, Goma *et al.* [18] formulated an optimisation model that takes advantage of the overlapping of home wireless LANs to aggregate the user traffic in as few GWs as possible. In both works, however, the possibility of transferring data among the GWs is not accounted for, nor is the caching capability of the GWs.

Han *et al.* [20] assumed a model which is quite close to the CDWMN one, and showed how such local connectivity and storage can be exploited to reduce the traffic on the access network. Yet, the energy issue is out of the scope of their work.

Valancius *et al.* [4] proposed NaDa, a distributed platform that uses ISP-controlled home gateways to provide computing and storage services to the end-users. NaDa adopts a peer-to-peer philosophy, but the GWs are not shared among the users and the connection among them is built and managed through the ISP infrastructure. Therefore, despite addressing the energy-saving topic, Valancius' work can hardly fit the CDWMN context. A very similar approach has been followed by Whiteaker *et al.* [21], who built a prototype of a service-hosting gateway that can be used by the ISP. Among the many services, it can cache data, but it is not devised to be shared among the users, nor to be connected to other gateways. The same can be said for the work of Den Hartog *et al.* [3], who assessed the pros and cons of caching various types of content in the GWs. The evaluation is in terms of bandwidth usage and blocking probability, without caring neither for the energy consumption nor for the availability of a WMN.

Manetti *et al.* [13] designed a centralised and a peer-to-peer architecture to deliver contents in community networks. The peer-to-peer architecture, in particular, can mimic the CDWMN model. However, the focus of the work is mostly on the protocol design and content indexing aspects, with no mention to the power consumption issues.

Alasaad *et al.* [14] studied the problem of energy consumption in sharing a given content over the CDWMN. In Alasaad's view, the CDWMN is at the same time origin and destination of a single "viral" content. Also, only some nodes can perform caching, and the content is retrieved by means of peer-to-peer procedures that are agnostic on the physical distance between the peers. Consequently, the achieved results cannot be applied to the CDWMN system under study in this paper.

Taghizadeh *et al.* [22] studied cooperative caching policies for reducing the provisioning cost in spontaneous networks formed by mobile devices. Although the optimisation model accounts for multi-hop paths, the energy consumption is not considered at all. Shevade *et al.* [23] devised an algorithm for placing the contents on the Access Points serving a vehicular network. The main focus of the authors is on content dissemination, based on the prediction (and exploitation) of the vehicle trajectories. No consideration is given to the energy topic.

Wang *et al.* [24] recently investigated a content-centric WMN, i.e. a WMN enhanced with a small set of communication relays and a subset of wireless mesh routers serving as storage nodes. Their goal was to optimally placing the communication relays in order to achieve the maximum throughput. Many points makes this work different from ours: not energy-efficient, no user assignment, storage on a few nodes only.

Finally, some research on home router sharing has been performed within the ongoing SmartenIT project [25]. In particular, Seufert *et al.* [26] proposed a framework that targets traffic offloading from mobile networks to WiFi, content prefetching and caching on the home routers, and content delivery. Router sharing and a trust mechanism are also part of the framework. The

[1] Residential community networks are access infrastructures in which the users build a (wireless) network among their home gateways in order to share the capabilities of the gateways and/or the contents hosted at each user's premises [19][20].

authors' main focus is on how information from online social networks and from user mobility can be exploited to prefetch and store popular contents on the home routers in order to reduce the load on the mobile networks. Energy reduction, however, is seen solely as a positive, but not quantified side effect. Also, the existence of direct wireless connections among the routers is not considered.

In summary, there is no previous study that faces the same problem as the one we present in the paper. We recall that our goal is to minimise the energy consumption by jointly considering and acting on the three aspects that are distinctive of the CDWMN: sharing the access to the GWs, sharing the cached contents, and transporting the contents among the GWs. Therefore, the cited works either address single aspects of the problem, or target substantially different systems.

As a final note, it is worth noting that the access networks currently contribute to a large portion of the energy consumed in telecommunication networks. This is confirmed by the increasing amount of research and standardisation activity aimed at improving the energy efficiency of the access technologies. As an example, the ADSL2 standard defines a low power state that modems can enter to save energy. The benefit provided by the adoption of sleep mode strategies in ADSL modems has been studied by Finamore et al. [27] by analysing the traffic features of a large set of ADSL lines. Furthermore, Bonetto et al. [28] characterised the energy consumption of Points of Presence (PoPs) on the basis of a large dataset of measurements collected from the network of a nation-wide Internet Service Provider in Italy. Then, they showed that sleep mode policies can be effectively implemented and allow reducing the energy consumption of ADSL modems with little or marginal impact on the Quality of Service offered to users. This result can indeed motivate and complement our investigation, since a smart user-GW association contributes to reducing the number of active GWs, and therefore increases the number of ADSL modems that can be put in sleep mode, thus incrementing the overall energy saving.

III. QETAC: QoS-based Energy and Traffic Allocation for CDWMN

A. Architecture of the CDWMN

The CDWMN system fits both the wired and wireless Internet service provider (ISP) scenarios. The former can take advantage of keeping control of the home routers / set-top boxes rented/leased to the users, thus deploying its content and energy management strategies to decrease its operational costs. As for the wireless ISP scenario, several applications cases can be found, such as municipal wireless, sparse campuses, open areas for sport and artistic events, rural communities, disaster relief.

The reference CDWMN architecture is shown in Fig. 1. The access gateways (GWs) are connected to the Internet by means of a high speed connection (e.g. Gigabit Ethernet, GPON, wireless point-to-point). The GWs can also be part of the CDN service, which decides the contents that are cached on each GW. Then, each GW is equipped with two different wireless interfaces. One interface is used to serve the assigned user terminals (UTs), while the other is used to connect with the neighbouring GWs. The two interfaces work on non-overlapping channels in order to avoid mutual interference. A UT can be assigned (associated) to one GW only. Typically, the UTs of a given user will be associated to the GW from which they receive the strongest signal. However, we do not prevent UTs from associating with a different GWs if a radio link is available and the optimum allocation requires so.

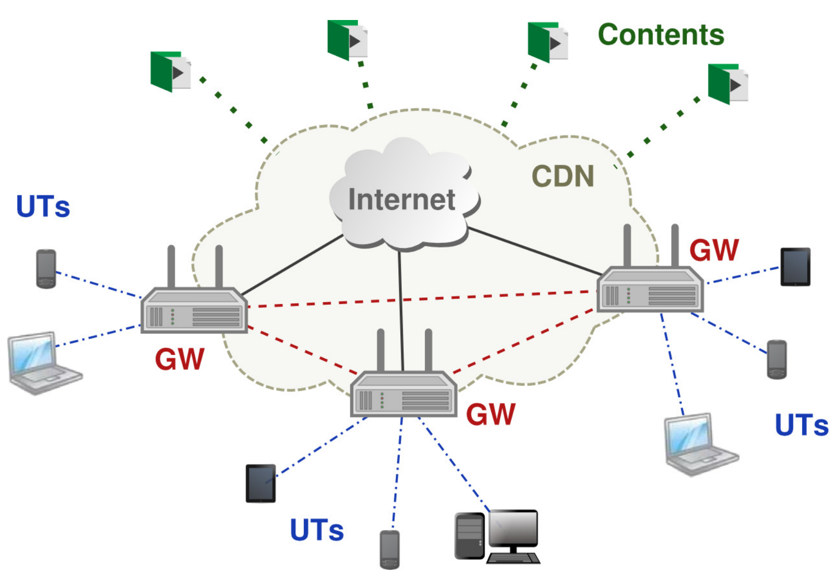

Fig. 1. Architecture of the CDWMN. Solid lines represent wired links, dashed lines wireless links, and dotted lines (towards the contents) virtual links.

Therefore, according to the illustrated CDWMN architecture, there are three possibilities for a GW to deliver a content to an assigned UT: (*i*) the content is cached on the GW itself, (*ii*) the content is cached on another GW that can be reached either directly or by means of a multi-hop connection, (*iii*) the content must be downloaded from the Internet.

B. The QETAC approach

QETAC exploits the distinctive features of the CDWMN to achieve energy efficiency and traffic offloading. Internet connectivity sharing, i.e. the possibility to associate any UT to any reachable GW, allows to power off some GWs. Caching on the GWs can be used to save Internet bandwidth (and energy). The mesh service enables the dissemination of the contents cached in any GW to any UT in the network.

To this purpose, we build a mathematical program that takes as input the placement of the contents on the GWs, the device connectivity, the user demands, and the QoS requirements, and decides (outputs) the UT-GW associations and the routing of the contents. Note that our focus is on already deployed WLANs, i.e. we do not solve the problem of choosing in which candidate sites the GWs shall be deployed. We assume this has been done in a previous phase, for example on the basis of the peak traffic demand pattern. Due to the already widespread adoption of WLANs, this hypothesis matches quite well with reality. Also, it allows to apply our method to existing networks, not just to the future ones. Similarly, our analysis starts from a given content placement pattern, laid out by a suitable CDN approach (see e.g. [4], [29], [30]). Therefore our work does not integrate the placement of the contents, but is complementary to the CDN deployment.

In the CDWMN abstraction used to formulate the mathematical model there are three kinds of nodes (UTs, GWs, and contents) and three kinds of links (GW-UT, GW-GW, and content-GW). The GW-UT and GW-GW are physical wireless links, and therefore are characterized by a data rate dependent on the modulation and radio propagation rules. Conversely, the content-GW links are logical, because we assume that every content on the Internet can be accessed by every GW with no data rate restrictions.

C. Notation

In the addressed problem we define the following sets:

- $\mathcal{G}$, the set of deployed gateways (GWs);
- $\mathcal{U}$, the set of user terminals (UTs);
- $\mathcal{C}$, the set of available contents (contents);
- $\mathcal{D} = \{(c,u) : u \in \mathcal{U}, c \in \mathcal{C}\}$, the set of content demands;
- $\mathcal{E}' = \{(g,u) : u \in \mathcal{U}, g \in \mathcal{G}, r_{gu} > 0\}$, the set of GW-UT edges;
- $\mathcal{E}'' = \{(g,h) : g,h \in \mathcal{G}, r_{gh} > 0\}$, the set of GW-GW edges;

where r_{gu} and r_{gh} are the data rates between a GW g and a UT u, and between a GW g and a GW h, respectively. Note that:

- the edges are directed, i.e. $(g,h) \neq (h,g)$,
- $\mathcal{E}'$ and $\mathcal{E}''$ are two disjoint sets, i.e. $\mathcal{E}' \cap \mathcal{E}'' = \emptyset$, and
- we have not made the set of content-GW edges explicit because it is not relevant to the model formulation.

The known parameters of the problem are:

- r_{ij} [b/s], the average data rate between a vertex $i \in \mathcal{G}$ and a vertex $j \in \{\mathcal{U} \bigcup \mathcal{G}\}$ – for example r_{gu} is the data rates between a GW g and a UT u;
- b_c [b/s], the rate needed for retrieving content c;
- d_U [s] and d_G [s], the delay bounds for each GW-UT link and for traversing the mesh of GWs, respectively;
- L [b], the average packet size, including all protocol headers;
- P^{GW} [W], the power consumption of a GW;
- E^I [J/b], the average energy for retrieving a bit from the Internet;
- E^W [J/b], the average energy for transmitting a bit through a wireless link (either GW-GW or GW-UT);
- t_{cg}, a binary flag that is set to 0 if content c is cached in GW g, to 1 otherwise.

We finally define the following binary variables:

- $q_g \in \{0,1\}$, which is set to 1 if GW g is powered on;
- $x_{gu} \in \{0,1\}$, which is set to 1 if UT u is assigned to GW g;
- $y^{cu}_{hg} \in \{0,1\}$, which is set to 1 if content c is delivered to UT u through link (h,g);
- $z^{cu}_{g} \in \{0,1\}$, which is set to 1 if UT u retrieves content c from GW g (i.e. either c is cached in g or c is downloaded from the Internet by g); note that u needs not to be assigned to g.

D. Mathematical programming model

The objective function of our problem is the combination of three elements. The first one, say P_{gw}, accounts for the power drained by the powered-on GWs:

$$P_{gw} = P^{GW} \sum_{g \in \mathcal{G}} q_g. \tag{1}$$

The second element, say P_{in}, represents the power consumption for transferring contents among the GWs:

$$P_{in} = E^W \sum_{(c,u)\in\mathcal{D}} \sum_{(h,g)\in\mathcal{E}''} (b_c\, y_{hg}^{cu}). \tag{2}$$

The third and last element, say P_{dl}, measures the traffic that is downloaded (not offloaded) from the Internet:

$$P_{dl} = E^I \sum_{g\in\mathcal{G}} \sum_{(c,u)\in\mathcal{D}} (t_{cg}\, z_g^{cu}\, b_c). \tag{3}$$

The three elements are merged into the single objective function:

$$\begin{aligned} \Omega &= (1-\gamma)\,(P_{gw}+P_{in}) + P_{dl} \\ &= (1-\gamma)\left\{P^{GW}\sum_{g\in\mathcal{G}} q_g + E^W \sum_{(h,g)\in\mathcal{E}''}\sum_{(c,u)\in\mathcal{D}} b_c\, y_{hg}^{cu}\right\} + E^I \sum_{g\in\mathcal{G}}\sum_{(c,u)\in\mathcal{D}} b_c\, t_{cg}\, z_g^{cu} \end{aligned} \tag{4}$$

Some aspects are worth noticing:

- Together, the terms P_{gw} and P_{in} give a measure of the power consumed by the elements and operations internal to the CDWMN, i.e. the GWs and the links among them. On the other hand, P_{dl} is proportional to the power consumed externally.
- A tuning parameter, $\gamma \in [0,1]$, weights the internal power terms P_{gw} and P_{in}, so that it is possible to assess the impact that various degrees of energy efficiency have on the global optimisation procedure. Specifically, when $\gamma = 1$ we perform a purely traffic oriented optimisation. Indeed, if $\gamma = 1$, (4) reduces to $\Omega = E^I \sum_{g\in\mathcal{G}} \sum_{(c,u)\in\mathcal{D}} (t_{cg}\, z_g^{cu}\, b_c)$, in which E^I is just a constant that has no impact on the optimisation, and thus the traffic only is accounted for. Conversely, if $\gamma = 0$, we have a pure energy efficient optimisation, because P_{gw}, P_{in} and P_{dl} are accounted on an equal basis, but traffic offloading is weighted in terms of the energy it consumes.
- P_{dl} is not weighted by parameter γ because we want to always account for the possibility of exploiting the cached contents (the reason is given in the next point). Therefore we impose that traffic offloading must always be part of the objective function. This also implies that we always account for the energy required by the Internet for bringing the data up to the GWs.
- If traffic offloading were not part of the objective function, the problem would turn into the greedy approach of minimising the energy consumed by the sole CDWMN. Such a problem has the trivial solution of turning off as many GWs as possible. In fact, when removing P_{dl} from (4), this would become:

  $$\Omega = P^{GW} \sum_{g\in\mathcal{G}} q_g + E^W \sum_{(h,g)\in\mathcal{E}''} \sum_{(c,u)\in\mathcal{D}} b_c\, y_{hg}^{cu}.$$

  Since downloading a content from the Internet has no penalty in the objective function, there is no advantage in moving the data over the GW-GW links. In other words, why should I move a content c from GW h to gw g (i.e. set $y_{hg}^{cu} = 1$), when I can download it for free at g?. Therefore the minimisation reduces, in practice, to the sole P_{gw}.

The objective of the mathematical program is to minimise Ω subject to:

$$x_{gu} + \sum_{(g,h)\in\mathcal{E}''} y_{gh}^{cu} - z_g^{cu} - \sum_{(h,g)\in\mathcal{E}''} y_{hg}^{cu} = 0 \qquad \forall (c,u)\in\mathcal{D}\ \forall (g,u)\in\mathcal{E}', \tag{5}$$

$$\sum_{(g,h)\in\mathcal{E}''} y_{gh}^{cu} - \sum_{(h,g)\in\mathcal{E}''} y_{hg}^{cu} - z_g^{cu} = 0 \qquad \forall (c,u)\in\mathcal{D}\ \forall (g,u)\notin\mathcal{E}', \tag{6}$$

$$\sum_{g\in\mathcal{G}} z_g^{cu} = 1 \qquad \forall (c,u)\in\mathcal{D}, \tag{7}$$

$$\sum_{(g,u)\in\mathcal{E}'} x_{gu} = 1 \qquad \forall u\in\mathcal{U}, \tag{8}$$

$$x_{gu} \le q_g \qquad \forall (g,u)\in\mathcal{E}', \tag{9}$$

$$y_{hg}^{cu} \le q_g\, q_h \qquad \forall (h,g)\in\mathcal{E}''\ \forall (c,u)\in\mathcal{D}, \tag{10}$$

$$z_g^{cu} \le q_g \qquad \forall g\in\mathcal{G}\ \forall (c,u)\in\mathcal{D}, \tag{11}$$

$$\sum_{(g,u)\in\mathcal{E}'} \frac{\sum_{(c,u)\in\mathcal{D}} b_c\, L}{2\, r_{gu}^2}\, x_{gu} \le d_U\,(1 - \sum_{(g,u)\in\mathcal{E}'} \frac{\sum_{(c,u)\in\mathcal{D}} b_c\, x_{gu}}{r_{gu}}) \qquad \forall g\in\mathcal{G}, \tag{12}$$

$$\sum_{(g,h)\in\mathcal{E}''} \frac{\sum_{(c,u)\in\mathcal{D}} b_c\, L}{2\, r_{gh}^2}\,(y_{hg}^{cu} + y_{gh}^{cu}) \le \frac{d_G}{H}\,(1 - \sum_{(g,h)\in\mathcal{E}''} \frac{\sum_{(c,u)\in\mathcal{D}} b_c\,(y_{hg}^{cu} + y_{gh}^{cu})}{r_{gh}}) \qquad \forall g\in\mathcal{G}. \tag{13}$$

Equations (5) and (6) are the flow conservation constraints (see below for a brief explanation), equations (7) impose that each UT u retrieves the content c from exactly one GW, i.e. that there is a single source for each demand (c,u), equations (8) impose that each UT must be assigned to exactly one GW, equations (9) and (10) impose that, if a GW is powered off, no UT or GW can be connected to it, equations (11) impose that no content can be retrieved from a powered-off GW, and equations (12) and (13) impose that the average waiting times in the GW-UT and GW-GW interfaces does not exceed a predefined bound. The next Section illustrates how (12) and (13) have been obtained.

Going back to the flow conservation constraints (5) and (6), a pictorial explanation is given in Fig. 2. Flow conservation is imposed in both cases at GW g. Note how we have employed the x_{gu} variable to also indicate that demand (c,u) is carried over the (g,u) link. In a more complete and formal version, we should have used some y_{gu}^{cu} variables, but it is also true that $x_{gu} = y_{gu}^{cu}\ \forall (g,u) \in \mathcal{E}'$, and thus we can simplify the model by using solely the x_{gu} variables. Also note that, since we do not perform demand splitting (i.e. no multipath routing), each link (g,h) shall either carry the whole demand or nothing. Thus, we can correctly formulate the flow conservation constraints by means of binary variables.

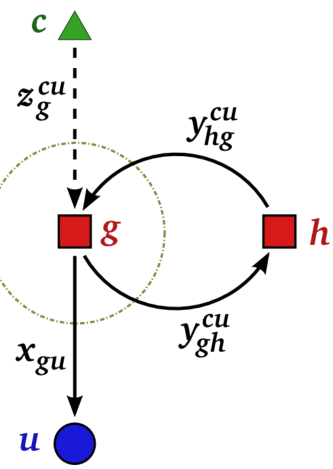

Fig. 2. Depiction of the flow conservation constraints (5). The same figure applies to (6) by removing user u, the (g,u) link, and the x_{gu} variable.

The above formulated model is an Integer Linear Programming (ILP) problem. Thus it can be solved by means of a Mixed-Integer Linear Programming solver, such as the IBM ILOG CPLEX Optimizer.

E. Delay bound implementation

In order to provide a reasonable limit for the flow delays, we took advantage of the queuing theory. In particular, we modelled the WLAN system as a network of M/D/1 queues with heterogeneous traffic classes/flows. Each queue represents a GW interface, in which the packet arrivals are exponential, with parameter λ_i, and the server implements a FCFS (first come, first serve) discipline with deterministic serving time $1/\mu_i$. For both parameters, i is the index of the traffic flow. Accordingly, the average waiting time in the queue is determined by the Pollaczek-Khinchine formula [31]:

$$E[T_W] = \frac{\sum_i \lambda_i E[T_{S,i}{}^2]}{2\,(1-\sum_i \rho_i)}, \tag{14}$$

where $\rho_i = \lambda_i/\mu_i$, and $T_{S,i} = 1/\mu_i$ is the serving time of class i. We can now express the general queuing parameters in terms of the parameters of our problem. Specifically, let $L_i = L\,\forall i$, and b_i the data rate of flow i (note that a flow does not necessarily coincide with a content demand). Then, the arrival rate λ_i is equal to b_i/L, the serving time μ_i is equal to r_i/L, and consequently $\rho_i = b_i/r_i$. Therefore, (14) becomes:

$$E[T_W] = \frac{\sum_i b_i L/r_i^2}{2\,(1-\sum_i b_i/r_i)}. \tag{15}$$

The last step is to convert and adapt (15) to fit our model. To this aim, we must separate the GW-UT links from the GW-GW links. For the former, we have that the flow index i corresponds to the GW-UT edge (g,u), in short $i \to (g,u)$. Accordingly, $r_i \to r_{gu}$. Then, the flow data rate b_i is built by the sum of all demands routed over the (g,u) link: $b_i \to \sum_{(c,u)\in\mathcal{D}} b_c$. Thus, operating the defined substitutions and imposing a delay bound of d_U, we obtain (12). Note that (12) also impose a capacity constraint, since $E[T_W]$ is bounded if and only if $\sum_i \rho_i < 1$, i.e. if the utilization of the queue is less than one. We obtain (13) in a similar fashion, with the sole difference that we must account for both directions of the (g,h) link and for the average number of hops H.

Note that, though this model is somewhat simplistic, it nevertheless allows us to describe the delay of the various flows in a reasonable way. This is sufficient for the purpose of our work, because our goal is assessing the energy-traffic optimisation tradeoff under some QoS constraints, whereas providing a thorough modelling of the WLAN system is out of the scope.

F. QETAC/TA – a version with traffic aggregation

In this subsection we formulate an alternative version of the QETAC approach that consists in aggregating the traffic carried along the GW-GW links (and downloaded from Internet) on the basis of the contents. With such aggregation we expect to achieve further energy and traffic savings, since we avoid traffic duplication. In addition, since less traffic is carried over the network, it is also easier to meet the delay bounds.

Firstly, we define the following further binary variables:

- $\vartheta_{hg}^{c} \in \{0,1\}$, which is set to 1 if content c is delivered through link (h,g);
- $\xi_{g}^{c} \in \{0,1\}$, which is set to 1 if content c is retrieved from GW g.

These new variables differ from y_{hg}^{cu} and z_{g}^{cu} in that they do not target the demands (c,u) but just the contents. The objective function for the QETAC/TA problem is also content-centric (rather than demand-centric) and is the following:

$$\Omega^{a} = (1-\gamma)P^{GW}\sum_{g\in\mathcal{G}} q_g + E^{I}\sum_{g\in\mathcal{G}}\sum_{c\in\mathcal{C}} b_c\, t_{cg}\, \xi_g^c + (1-\gamma)E^{W}\sum_{(h,g)\in\mathcal{E}''}\sum_{c\in\mathcal{C}} b_c\, \vartheta_{hg}^{c} \tag{16}$$

The problem maintains unaltered the constraints (5), (6), (7), (8), (9), and (12), whereas (10), (11), and (13) change as follows:

$$\vartheta_{hg}^{c} \le q_g\, q_h \qquad \forall (h,g)\in\mathcal{E}''\ \forall c\in\mathcal{C}, \tag{17}$$

$$\xi_{g}^{c} \le q_g \qquad \forall g\in\mathcal{G}\ \forall c\in\mathcal{C}, \tag{18}$$

$$\sum_{(g,h)\in\mathcal{E}''} \frac{\sum_{(c,u)\in\mathcal{D}} b_c\, L}{2\, r_{gh}^{2}}\,(\vartheta_{hg}^{c}+\vartheta_{gh}^{c}) \le \frac{d_G}{H}\,(1-\sum_{(g,h)\in\mathcal{E}''}\frac{\sum_{(c,u)\in\mathcal{D}} b_c\,(\vartheta_{hg}^{c}+\vartheta_{gh}^{c})}{r_{gh}}) \qquad \forall g\in\mathcal{G}. \tag{19}$$

Finally, the following two sets of constraints are necessary to bind the content-based variables to the demand-based variables:

$$y_{hg}^{cu} \le \vartheta_{hg}^{c} \qquad \forall (h,g)\in\mathcal{E}''\ \forall (c,u)\in\mathcal{D}, \tag{20}$$

$$z_{g}^{cu} \le \xi_{g}^{c} \qquad \forall g\in\mathcal{G}\ \forall (c,u)\in\mathcal{D}. \tag{21}$$

IV. Performance evaluation

A. Scenario and parameter values

The performance of QETAC has been tested over a series of 11 network scenarios. For each scenario, we generated and solved twenty instances. The results have been averaged over the whole set of instances and scenarios.

The scenarios are characterised by a number of input parameters, some of them taking different values in each scenario. Table I reports the values of the varying parameters. It comprises the dimension of the sets $\mathcal{G}$, $\mathcal{U}$, and $\mathcal{C}$, the cache size (number of contents) on each GW (S), and the popularity exponent of the content demand (α). The "standard" scenario, with all parameter values taken from the central column, is used as the reference one. Starting from it, we have changed one parameter value per scenario.

TABLE I
Parameter values for the tested scenarios.

Parameter	Minimum	Standard	Maximum
$\lvert\mathcal{G}\rvert$	10	30	50
$\lvert\mathcal{U}\rvert/\lvert\mathcal{G}\rvert$	2	4	6
$\lvert\mathcal{C}\rvert/\lvert\mathcal{U}\rvert$	100	1000	10000
$S/\lvert\mathcal{U}\rvert$	0.25	2.5	25
α	0.5	0.65	0.8

GWs and UTs have been placed in a fictitious test area of varying size in order to keep the GW density constant. To avoid heavily unbalanced instances, the test field has been divided into a regular grid and the GWs and UTs have been evenly distributed across the grid squares. In each instance, however, the positions of the GWs and UTs have been randomly determined (within the assigned square).

The energy consumed by every GW has been set to P^{GW} = 15 W, which is a typical value for enterprise wireless routers. The energy consumption for retrieving a bit of content via Internet has been set to E^{I} = 39 μJ/b [32], and the energy consumption for transferring a bit of content over a wireless link to E^{W} = 0.02 μJ/b [33].

To determine the GW-GW and GW-UT data rates (i.e. the various r_{gu} and r_{gh}), we employed a simplified version of the COST-231 path loss model [34], which allows to account for various propagation aspects, such as the presence of walls and other obstacles, and the use of a realistic path loss exponent. The values of all parameters have been extracted from real measurements [35], [36] and data sheets (such as [37]). Then, the data rates have been extracted from the computed signal-to-noise ratio (σ_{ij}, with $i\in\mathcal{G}$ and $j\in\{\mathcal{U}\bigcup\mathcal{G}\}$) according to:

$$r_{ij} = \min\{\psi_{ij}(\sigma_{ij}), r_{max}\}, \tag{22}$$

where $\psi_{ij}(\sigma_{ij})$ is a proper function that might depend on several parameters, such as the specific modulation and coding schemes and the medium access overhead – in our case we referred to the work of Zhang *et al.* [38], and r_{max} is used to cap r_{ij} to the maximum rate achievable by the physical link, say r_{max} = 54 Mbps (as per the IEEE 802.11a/g).

The content demand has been modelled according to a content popularity that follows the widely adopted Zipf distribution [39], with the popularity exponent α varying from 0.5 to 0.8 (as shown in Table I) based on the empirical observation reported in [40], [41].

The cached contents in the GWs have been modelled by means of the same Zipf distribution. We reckon that this is a sensible behaviour for a "smart" caching strategy, which would presumably store with the highest probability the most popular contents [39]. For each GW we assumed a number of cached contents equal to S (the system is in the steady state). We have considered a somewhat cooperative caching strategy, in which the contents to be cached are selected so as the same contents are not cached in adjacent GWs.

Finally, based on [42], we have considered the discrete distribution of the data rate of the contents (b_c) shown in Fig. 3.

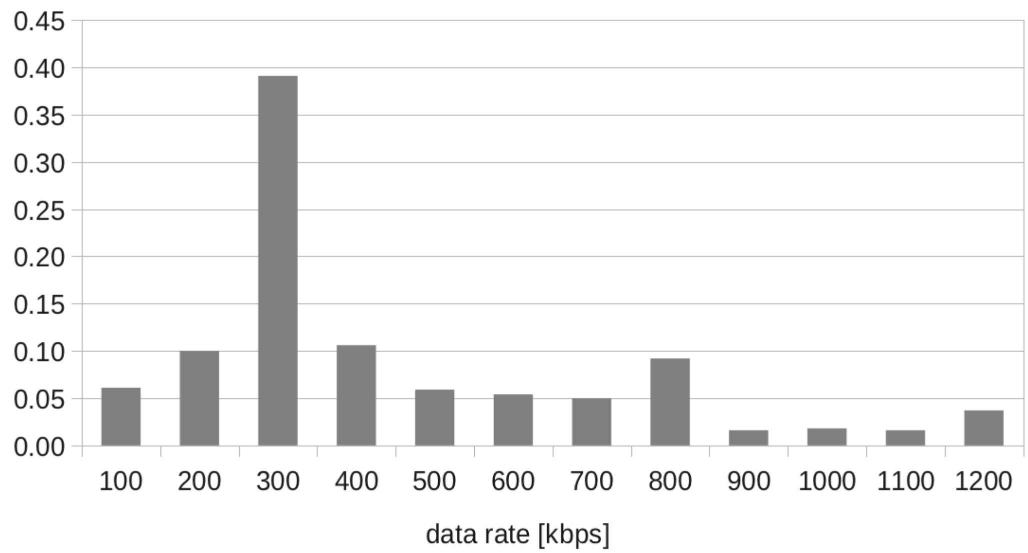

Fig. 3. YouTube video encoding rate distribution.

As for the delay aspect, we set d_U = 50 ms and d_G = 50 ms. Consequently, the delay that is introduced by the CDWMN should be, at most, in the order of 100 ms, which is generally suitable for most multimedia contents. For the H parameter we have obtained an upper bound by performing the same experiments described in Section IV-B using the capacity constraints (24)-(25) (see below) in place of the delay bounds (12)-(13), and then by extracting the 99-percentile of the average hop count. The resulting value is $H = 1.095$.

B. Reference benchmarks

The performance of our model is assessed in terms of several metrics. For most of them we adopted a common benchmark, which is the currently most widespread connection model in enterprise and residential WLANs. In short, the GWs form separate connectivity islands, do not perform content caching, are not shared nor connected with each other, and are always active. The UTs simply associate to the GW from which they receive the strongest signal. In this case, the power consumption of this network model, named SCI (Single Connectivity Islands) for convenience, can be readily computed as:

$$P_{\text{SCI}} = |\mathcal{G}| \cdot P^{GW} + E^I \sum\nolimits_{(c,u)\in\mathcal{D}} b_c. \tag{23}$$

Two further benchmarks are built by replacing (12) and (13) with the classical capacity constraints:

$$\sum_{(g,u)\in\mathcal{E}'} \left\{ \frac{\sum_{(c,u)\in\mathcal{D}} b_c}{r_{gu}} \cdot x_{gu} \right\} \leq q_g \qquad \forall g \in \mathcal{G}, \tag{24}$$

$$\sum_{(h,g)\in\mathcal{E}''} \sum_{(c,u)\in\mathcal{D}} \left\{ \frac{b_c}{r_{hg}} \cdot (y^{cu}_{hg} + y^{cu}_{gh}) \right\} \leq q_g \qquad \forall g \in \mathcal{G}, \tag{25}$$

and by setting γ to either 1 or 0, in order to perform, respectively, pure energy efficiency (EE) or traffic offloading (TO) optimisation. The resulting problems, which have no QoS guarantees, yield an upper bound on the energy or traffic performance of the system. However, as it will be shown later on, the obtained solutions are hardly of any practical use because of the unbounded delay.

C. Results

The most prominent measure of energy efficiency is the amount of consumed power. Fig. 4 depicts the power consumption of the QETAC approach for various values of γ – indicated as QETAC(γ). EE and TO have also been reported. All values are normalised to the power consumption of SCI.

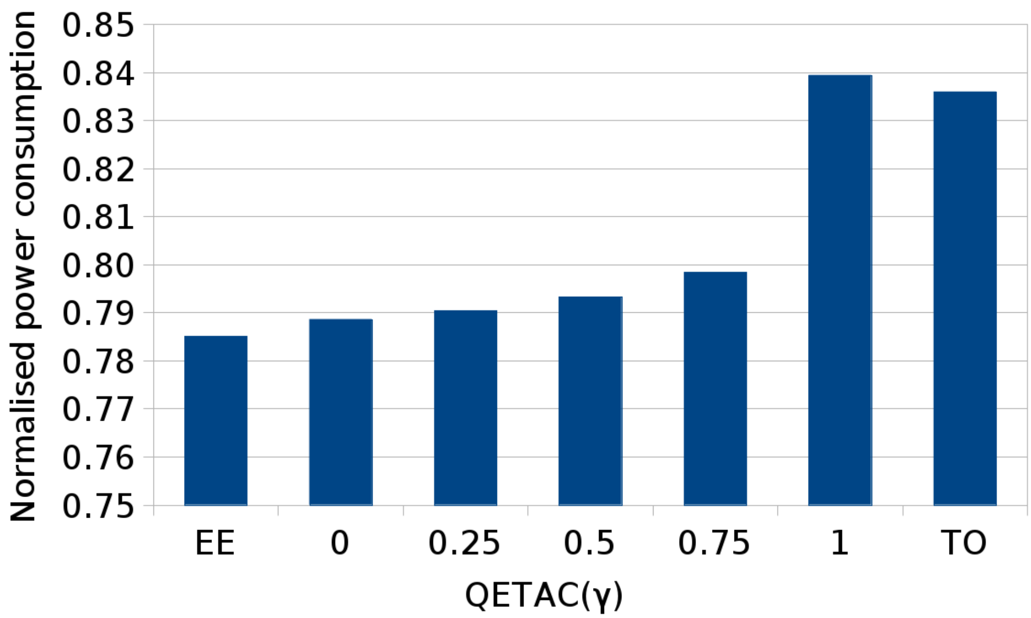

Fig. 4. Normalised power consumption of QETAC as a function of the γ value. Lower is better.

At first, we can see that in all cases QETAC provides a fair amount of power saving with respect to SCI. This amount, however, is variable, as the traffic oriented version provides quite a small energy efficiency improvement. More in detail, QETAC(0) can save up to 21.1% of power, which gradually reduces as γ increases, touching the minimum (16.1%) for $\gamma = 1$. From a comparison with the non-QoS versions, it emerges that the efficiency loss of QETAC is very limited. EE yields a 21.5% saving, and TO 16.4%. Thus, the energy loss due to the delay constraints is indeed minimal.

Fig. 5 shows the outcome of the experiments from the traffic offloading point of view. In this case the best result is, obviously, achieved by QETAC(1), which allows to save up to 19.7% of Internet traffic. A gradual decrease goes along with the reduction of γ, touching a minimum for $\gamma = 0$ (17.7%). However, the distance between the two extremes is quite limited, so that we can affirm that all versions of QETAC(γ) yield satisfying results. Indeed, though EE and TO performs better than QETAC(0) and QETAC(1), the gap is almost negligible (about 0.5%).

The third parameter we examine is the delay d experimented by the demands (in all scenarios and all instances). The delay is computed, for each demand, as the sum of all waiting times in the queues plus all transmission times. Table II reports four statistics on the delay observation set: the mean value ($\overline{d}$), the 99-percentile, and the percentage of observed values greater than the bounds "imposed" by (12) and (13). Incidentally, we recall that (12)-(13) impose a limit on the *average* queuing times of the demands, which is a statistical property of an aleatory variable of which we are now observing the realisations. Therefore it may occur that $d > d_U$ or even $d > d_U + d_G$, but their occurrence should be kept to acceptable levels.

TABLE II
DELAY FIGURES AND FRACTION OF POWERED-OFF GWS FOR QETAC, EE, AND TO.

	EE	QETAC(γ)					TO
		0	0.25	0.5	0.75	1	
Mean delay value, $\overline{d}$ [ms]	451	15.2	15.4	15.3	15.7	16.5	1519
99-percentile of d [ms]	2401	97.5	98.7	106	110	112	2153
$d > 50$ms [%]	26.4	6.1	6.4	6.7	7.1	7.9	24.0
$d > 100$ms [%]	14.1	0.93	0.96	1.20	1.30	1.46	15.7
Powered-off GWs [%]	34.6	34.2	29.0	25.5	21.5	0	0

As already anticipated, the delay figures of EE and TO do not allow for a plain transposition of the provided resource allocation into the real world. In both cases, $\overline{d}$ is far beyond the QoS requirement, with a conspicuous number of demands exceedings both QoS bounds. Conversely, QETAC promises much lower delays, with about one hundredth of the demands failing to achieve the QoS objective of 100 ms.

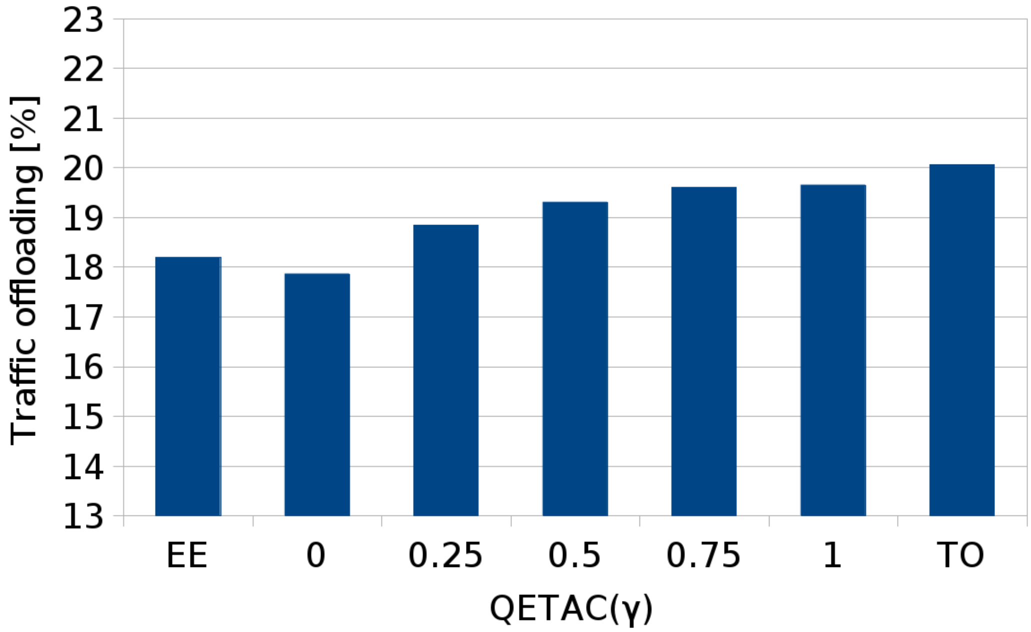

Fig. 5. Percentage of traffic offloading as a function of the γ value. Higher is better.

By means of the cumulative distribution function (CDF), illustrated in Fig. 6, we can complete the picture about the delay performance. The behaviour is very similar across all γ values, with just QETAC(1) having a slightly more gradual trend. For all of them, the knee is roughly at 50 ms. The performance difference with either EE or TO is apparent.

Further insights can be obtained from the percentage of GWs that are powered off by the various models (last row of Table II). It emerges that QETAC(1) and TO do not disactivate any GW. This is the obvious consequence of having no GW component in the objective function. Since both approaches focus exclusively on maximing the offloaded traffic, turning off some GWs might be inconvenient. The other interesting result is about EE and QETAC(0). Both approaches target energy efficiency only, and they yield (almost) the same highest fraction of inactive GWs. However, QETAC(0) is subject to the more stringent delay bounds. Therefore it appears that meeting the QoS parameter can be achieved without loosing much energy efficiency. For the other values of γ, QETAC gradually increases the number of active GWs in order to allow for more traffic to be carried over the CDWMN. Obviously, this leads to less power saving, as already shown in Fig. 4.

A further metric we analyse is the hit rate (η), which measures the capability to find and retrieve the demanded contents from within the CDWMN. The hit rate is defined as:

$$\eta = \frac{\sum_{(c,u)\in\mathcal{D}} \sum_{g\in\mathcal{G}} (1-t_{cg}) \cdot \bar{z}_g^{cu}}{|\mathcal{D}|} \cdot 100. \tag{26}$$

Fig. 7 plots the bars for all approaches. A distinction is performed on the basis of the location of the content. With the term "local" we refer to contents cached on the GW a UT is currently assigned to ($\bar{x}_{gu} = 1$), whereas "remote" indicates contents cached on other GWs in the CDWMN. This implies that a remote content must be retrieved by means of a multi-hop path, and might incur a higher delay (subject to a $d_G + d_U$ bound) with respect to a local content (whose delay bound is solely d_U).

As expected, the greater the γ, the more the objective function is traffic oriented, and the highest is the achieved hit rate. The biggest jump is between QETAC(0) and QETAC(0.25), suggesting that even a moderate lean towards the traffic offloading goal is enough to sensibly improve the performance of the optimisation model. Also worth noting is the fact that the local hit rate grows more slowly than the remote one. This suggests that increasing γ pushes for a higher utilisation of the mesh facility, and therefore tends to activate more GWs in order to support the delivery along the delay-bounded paths. Lastly, note how EE and TO does not get sensibly higher η values. Therefore, QETAC makes it feasible to achieve both almost optimal energy/traffic savings without sacrifying the service quality.

In terms of computational complexity, Fig. 8 shows the average CPU time (in seconds) for solving the various optimisation problems on a PC equipped with a 2.27 GHz 64-bit processor. The differences among the approaches are indeed conspicuous. Remarkably, the introduction of the delay constraints improves the solving time, as it can be inferred by the gaps between EE and QETAC(0), and TO and QETAC(1), respectively. Also, energy-efficiency requires much more computational resources than traffic offloading. Therefore there is a gradual decrease in the solving times as γ goes from 0 to 1, the exceptions being QETAC(0.5) and QETAC(1). The former pays the fact of not having a dominating term, i.e. it must weigh the two objectives in exactly the same manner, whereas the other QETAC variants benefits from having a stronger component that partly drives

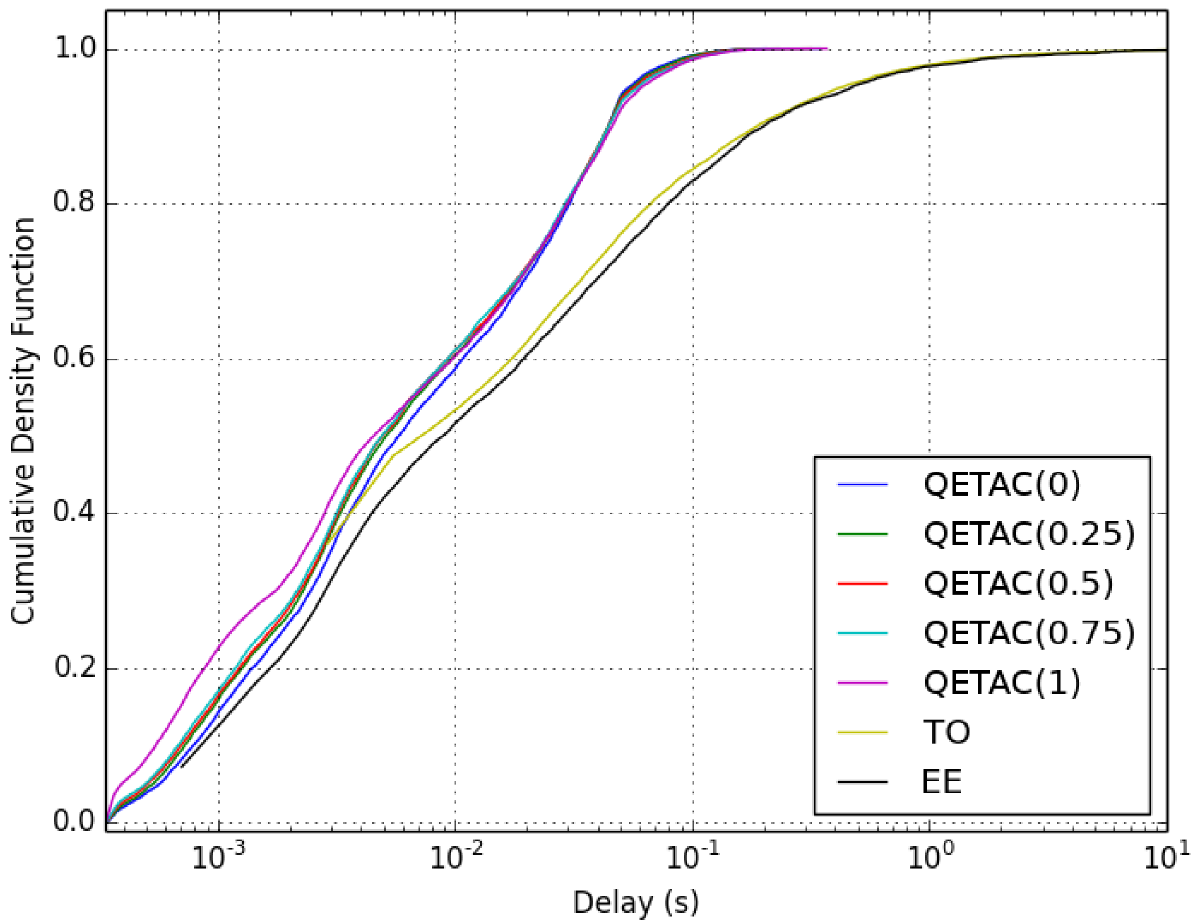

Fig. 6. Cumulative distribution function of QETAC(γ). Note that the x-axis is in logarithmic scale.

the objective function. At the extreme end, QETAC(1) takes advantage of having solely the traffic-oriented goal, which is much simpler to achieve.

Finally, in this last part of the results section, we provide an insight on the performance of QETAC/TA. We have chosen to compare it with QETAC on the basis of a single value for γ, in order to keep the analysis flowing. Table III summarises the differences between the two approaches on the basis of $\gamma = 0.5$.

TABLE III
PERFORMANCE FIGURES FOR QETAC AND QETAC/TA FOR $\gamma = 0.5$.

Parameter	QETAC	QETAC/TA
Normalised power consumption	0.793	0.793
Traffic offloading	19.31	19.27
Local hit rate	7.21	7.19
Remote hit rate	13.46	12.98
Powered-off GWs [%]	25.5	25.6
Mean delay value, $\overline{d}$ [ms]	18.1	18.8
99-percentile of d	105.6	106.5
$d >$ 50ms [%]	6.67	6.90
$d >$ 100ms [%]	1.20	1.20
Solving time [s]	1188	1239

For most metrics, the two approaches behaves in a very similar manner. They provide the same amount of power saving, and almost the same traffic offlaoding, number of inactive GWs, and local hit rate percentage. A more meaningful difference is in the remote hit rate, which is smaller for QETAC/TA, because it tends to trade off some opportunities of fetching a content from remote GWs with the possibility of saving energy and traffic by means of the aggregation feature.

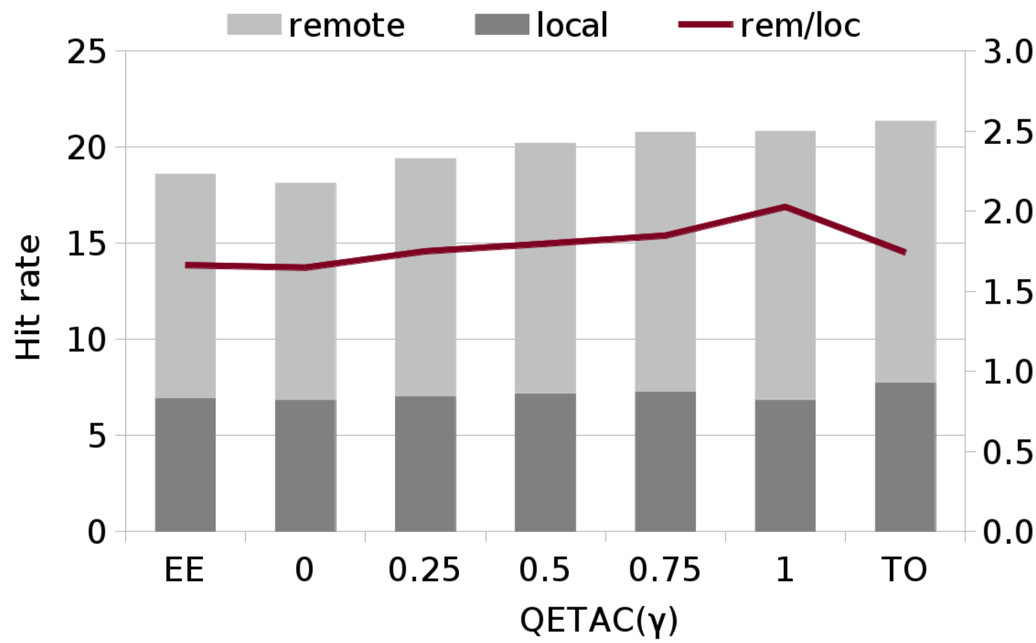

Fig. 7. Local and remote hit rates (stacked bars) plus remote/local hit rate ratio, as a function of the γ value.

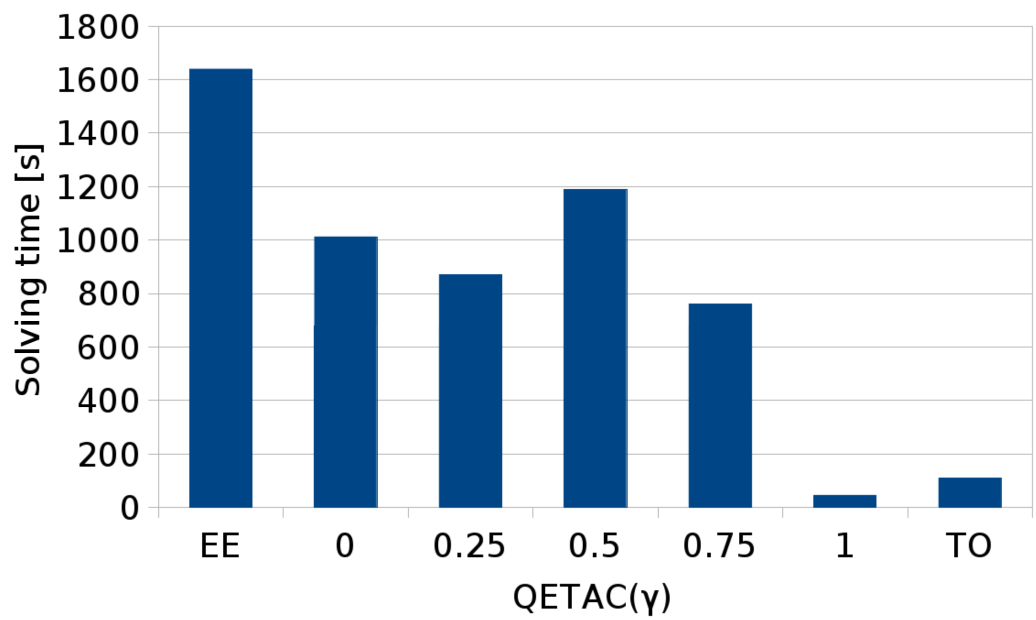

Fig. 8. Solving times as a function of the γ value.

In terms of delay, the performance of QETAC/TA is still very close to that of QETAC. The delays are slightly larger, but on the whole the QoS requirement is kept at the same level, especially when looking at the number of demands whose delay exceeds the 100 ms bound. The solving times are also comparable.

V. Conclusion

The paper presented and discussed QETAC, a method for optimising the energy efficiency and traffic offloading performance of a wireless access network with caching and mesh capabilities under quality of service requirements. From a large set of experiments, based on practical rate and consumption figures, it emerged that QETAC is capable of providing a sensible saving in both objectives. Furthermore, combining the two goals into a single objective function, allows for the flexible tuning of the operating point of QETAC. We also proved that it is possible to integrate bandwidth requirements and realistic delay bounds in the optimisation model with the twofold advantage of offering an acceptable level of service to the end users without producing any meaningful loss in neither energy efficiency nor traffic offloading. The combination of these two achievements makes the allocation optimal not just from a theoretical point of view, but also deployable in practical scenarios. Finally, we have also assessed the possible advantages of performing traffic aggregation. The results, however, did not highlight any meaningful improvement with respect to the "basic" QETAC approach.

In summary, the QETAC approach can indeed be profitable for network and service operators, which would have an efficient and viable tool for quantifying and choosing the energy/traffic tradeoff point for their networks. In addition, it can be integrated with energy-saving techniques at the ISP points-of-presence, thus increasing the overall energy-efficient of the access network.

Acknowledgment

This work was partially supported by the Italian Ministry of University and Research under the "GreenNet" FIRB project.

References

[1] Pallis G, Vakali A. Insight and perspectives for content delivery networks. *Communications of the ACM* Jan 2006; **49**(1):101–106, doi:10.1145/1107458.1107462.

[2] Spagna S, Liebsch M, Baldessari R, Niccolini S, Schmid S, Garroppo R, Ozawa K, Awano J. Design principles of an operator-owned highly distributed content delivery network. *IEEE Communications Magazine* 2013; **51**(4):132–140, doi:10.1109/MCOM.2013.6495772.

[3] den Hartog F, Bastiaans B, Blom M, Pluijmaekers M, van der Mei R. The use of Residential Gateways in Content Delivery Networking. *Proc. Australian Telecommunication Networks and Applications Conference (ATNAC)*, Sydney, 2004.

[4] Valancius V, Laoutaris N, Massoulié L, Diot C, Rodriguez P. Greening the internet with nano data centers. *CoNEXT*, 2009; 37–48, doi:10.1145/1658939.1658944.

[5] Cheng T, Wu J, Yang F, Leu J. IEEE 802.16e/m energy-efficient sleep-mode operation with delay limitation in multibroadcast services. *International Journal of Communication Systems* 2014; **27**(1):45–67, doi:10.1002/dac.2342.

[6] Li Z, Jiang H, Pan Z, Liu N, You X. Energy spectral efficiency tradeoff in downlink OFDMA network. *International Journal of Communication Systems* 2014; :in pressdoi:10.1002/dac.2725.

[7] Lorincz J, Capone A, Begusic D. Optimized network management for energy savings of wireless access networks. *Computer Networks* Feb 2011; **55**(3):514–540, doi:10.1016/j.comnet.2010.09.013.

[8] Couto da Silva AP, Meo M, Marsan MA. Energy-performance trade-off in dense WLANs: A queuing study. *Computer Networks* 2012; **56**(10):2522–2537, doi:10.1016/j.comnet.2012.03.017.

[9] Hashimoto M, Hasegawa G, Murata M. An analysis of energy consumption for TCP data transfer with burst transmission over a wireless LAN. *International Journal of Communication Systems* 2014; :in pressdoi:10.1002/dac.2832.

[10] De La Oliva A, Banchs A, Serrano P. Throughput and energy-aware routing for 802.11 based mesh networks. *Computer Communications* 2012; **35**(12):1433–1446, doi:10.1016/j.comcom.2012.04.004.

[11] Oliveira LML, de Sousa AF, Rodrigues JJPC. Routing and mobility approaches in IPv6 over LoWPAN mesh networks. *International Journal of Communication Systems* 2011; **24**(11):1445–1466, doi:10.1002/dac.1228.

[12] Xue L, Guan X, Liu Z, Yang B. TREE: Routing strategy with guarantee of QoS for industrial wireless sensor networks. *International Journal of Communication Systems* 2014; **27**(3):459–481, doi:10.1002/dac.2376.

[13] Manetti V, Canonico R, De Donato W, Ventre G, Mauthe A, Tyson G. Next Generation CDN Services for Community Networks. *NGMAST*, 2009; 89–94, doi:10.1109/NGMAST.2009.45.

[14] Alasaad A, Gopalakrishnan S, Nicanfar H, Leung V. Green content distribution in Wireless Mesh Networks with infrastructure support. *ICC*, 2012; 5896–5900, doi:10.1109/ICC.2012.6364736.

[15] Mor V, Kumar H. Energy efficient techniques in Wireless Mesh Network. *Recent Advances in Engineering and Computational Sciences (RAECS)*, IEEE, 2014.

[16] Jiang W, Ioannidis S, Massoulié L, Picconi F. Orchestrating massively distributed CDNs. *Proc. 8th international conference on Emerging networking experiments and technologies*, ACM, 2012; 133–144.

[17] Rossi C, Casetti C, Chiasserini CF. An Energy Efficient Protocol for Gateway-Centric Federated Residential Access Networks 2011. URL http://arxiv.org/abs/1105.3023, computing Research Repository (CoRR).

[18] Goma E, Canini M, Lopez Toledo A, Laoutaris N, Kostić D, Rodriguez P, Stanojević R, Yagüe Valentin P. Insomnia in the access: or how to curb access network related energy consumption. *Computer Communication Review* August 2011; **41**(4):338–349, doi:10.1145/2043164.2018475.

[19] Szabó C, Farkas K, Horváth Z. Motivations, design and business models of wireless community networks. *Mobile Networks and Applications* April 2008; **13**(1-2):147–159, doi:10.1007/s11036-008-0031-y.

[20] Han D, Andersen D, Kaminsky M, Papagiannaki D, Seshan S. Hulu in the neighborhood. *Third International Conference on Communication Systems and Networks (COMSNETS)*, 2011, doi:10.1109/COMSNETS.2011.5716501.

[21] Whiteaker J, Schneider F, Teixeira R, Diot C, Soule A, Picconi F, May M. Expanding home services with advanced gateways. *Computer Communication Review* Oct 2012; **42**(5):37–43, doi:10.1145/2378956.2378962.

[22] Taghizadeh M, Micinski K, Biswas S, Ofria C, Torng E. Distributed Cooperative Caching in Social Wireless Networks. *IEEE Transactions on Mobile Computing* 2013; **12**(6):1037–1053, doi:http://doi.ieeecomputersociety.org/10.1109/TMC.2012.66.

[23] Shevade U, Chen YC, Qiu L, Zhang Y, Chandar V, Han MK, Song HH, Seung Y. Enabling high-bandwidth vehicular content distribution. *Proc. 6th International Conference on emerging Networking EXperiments and Technologies (CoNEXT)*, ACM: New York, NY, USA, 2010; 23:1–23:12, doi:10.1145/1921168.1921199.

[24] Wang J, Lu K, Zhang S, Fan J, Zhu Y, Cheng B. An efficient communication relay placement algorithm for content-centric wireless mesh networks. *International Journal of Communication Systems* 2013; :n/adoi:10.1002/dac.2666. In press.

[25] Stiller B, Hausheer D, Hoßfeld T. Towards a Socially-Aware Management of New Overlay Application Traffic Combined with Energy Efficiency in the Internet (SmartenIT). *The Future Internet*, *Lecture Notes in Computer Science*, vol. 7858, Galis A, Gavras A (eds.). Springer Berlin Heidelberg, 2013; 3–15, doi:10.1007/978-3-642-38082-2_1. URL http://dx.doi.org/10.1007/978-3-642-38082-2_1.

[26] Seufert M, Burger V, Hossfeld T. HORST - Home router sharing based on trust. *Network and Service Management (CNSM), 2013 9th International Conference on*, 2013; 402–405, doi:10.1109/CNSM.2013.6727865.

[27] Bonetto E, Finamore A, Munafo MM, Fiandra R. Sleep Mode at the Edge: How Much Room is There? *Proc. of Networks* 2012; .

[28] Bonetto E, Finamore A, Mellia M, Fiandra R. Energy efficiency in access and aggregation networks: From current traffic to potential savings. *Computer Networks* 2014; **65**:151–166.

[29] Borst S, Gupta V, Walid A. Distributed Caching Algorithms for Content Distribution Networks. *Proc. IEEE INFOCOM*, 2010, doi:10.1109/INFCOM.2010.5461964.

[30] Al-Arnaout Z, Fu Q, Frean M. A divide-and-conquer approach for content replication in WMNs. *Computer Networks* 2013; doi:10.1016/j.comnet.2013.09.016. In press (DOI 10.1016/j.comnet.2013.09.016).

[31] Kleinrock L. *Queueing Systems*, vol. II: Computer Applications. Wiley Interscience, 1976.
[32] Baliga J, Ayre R, Hinton K, Sorin WV, Tucker RS. Energy Consumption in Optical IP Networks. *Journal of Lightwave Technology* Jul 2009; **27**(13):2391–2403.
[33] Halperin D, Greenstein B, Sheth A, Wetherall D. Demystifying 802.11n power consumption. *Proc. International conference on Power aware computing and systems*, USENIX Association, 2010.
[34] European Comission. COST 231 – Digital Mobile Radio Towards Future Generations Systems. Final Report 1999.
[35] Wilson R. Propagation Losses Through Common Building Materials: 2.4 GHz vs 5 GHz. *Technical Report E10589*, Magis Networks, Inc. August 2002.
[36] de Francisco R. Indoor Channel Measurements and Models at 2.4 GHz in a Hospital. *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, 2010; 1–6, doi:10.1109/GLOCOM.2010.5683218.
[37] Cisco. Aironet 3600 Series Access Point. URL http://www.cisco.com/en/US/products/ps11983/index.html.
[38] Zhang J, Tan K, Zhao J, Wu H, Zhang Y. A Practical SNR-Guided Rate Adaptation. *Proc. IEEE Conference on Computer Communications (INFOCOM)*, 2008; 2083–2091, doi:10.1109/INFOCOM.2008.274.
[39] Haßlinger G, Hartleb F. Content delivery and caching from a network provider's perspective. *Computer Networks* 2011; **55**(18):3991–4006, doi:http://dx.doi.org/10.1016/j.comnet.2011.07.026.
[40] Gill P, Arlitt M, Li Z, Mahanti A. Youtube traffic characterization: a view from the edge. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, 2007.
[41] Cheng X, Dale C, Liu J. Statistics and Social Network of YouTube Videos. *IWQoS*, 2008.
[42] Ameigeiras P, Ramos-Munoz JJ, Navarro-Ortiz J, Lopez-Soler J. Analysis and modelling of YouTube traffic. *Transactions on Emerging Telecommunications Technologies* 2012; **23**(4):360–377, doi:10.1002/ett.2546.