Evaluating User-Centric Multihomed Flow Management in Multi-User Scenarios

Xi Li*, Olivier Mehani†, Ramón Agüero‡, Umar Toseef*, Yasir Zaki*, Carmelita Görg*
*University of Bremen, Bremen, Germany, Email: xili,umr,yzaki,cg@comnets.uni-bremen.de
†Nicta, Eveleigh, Sydney, NSW, Australia, Email: olivier.mehani@nicta.com.au
‡University of Cantabria, Santander, Spain, Email: ramon@tlmat.unican.es

***Abstract*—Modern mobile devices comprise multiple interfaces for heterogeneous network technologies. However, currently implemented mechanisms to decide which one(s) to use and distribute application flows accordingly (*i.e.*, solving the multihomed flow management problem, MFM) are rather coarse and do not leverage the opportunities. A user-centric quality-aware (QA) MFM has been proposed which optimises network use based on user-perceivable metrics such as application quality as well as energy and monetary costs. This paper refines this approach by providing a single method for both real-time and elastic (*i.e.*, TCP-based) traffic, and uses realistic available capacity estimation. We evaluate this method in OPNET-simulated LTE and WLAN mobile networks. We study the impact of methods used to trigger the decision algorithm, and investigate the influence of an increasing number of users employing the QA-MFM technique on both the user-perceivable metrics and the global network performance. We find that on-demand triggering performs better than a static periodic method. We also demonstrate that the proposed approach out-performs classical network selection techniques in terms of application quality. We also show that the QA-MFM is not too greedy as to not scale with a number of users, and has a positive effect on the network loads, by preemptively adapting applications parameters to match network conditions.**

***Index Terms*—Network management, Algorithm/protocol design and analysis, Mobile communication systems, Simulation**

I. Introduction

Mobile user devices, as well as forthcoming 4G standards, support multiple wireless network technologies (*e.g.*, WLAN or 3G/LTE) with different characteristics. This opens up the opportunity to distribute traffic over the available networks in order to realise “Always Best Connected” network access [1]. However, to date, only the simplest methods (“only use cellular connectivity if WLAN is unavailable”) are really deployed in user devices [2].

More advanced techniques are proposed in the literature, and focus mainly on two orthogonal axes: access network selection and scheduling of application flows. Another aspect, application parameters, is also sometimes included, usually in the form of cross-layer designs. It was however argued that these three aspects should be considered as part of the single *Multihomed Flow Management* (MFM) problem [3]. This was further supported by propesing to optimise those metrics which are really relevant to the application or its user. Initial simulation results demonstrated the validity of this Quality-Aware Multihomed Flow Management (QA-MFM) in comparison to more classical techniques [4]. The conclusions were however limited due to some of the assumptions taken (single user, single application and ideal capacity estimation) and, while encouraging, required a more thorough study.

In this paper, we lift these assumptions, introduce improvement to the decision algorithm algorithm (real- and non real-time traffic, and flow redistribution periodicity) and provide more insight into the performance of QA-MFM, both from a user and a network operator’s point of view. We first generalise the binary integer programming (BIP) formulation to allow for the management of both real-time and elastic flows in parallel. We extend the simulations from one to multiple users running the decision algorithm in parallel, and consequently update realistic available-capacity estimation methods. We also consider triggering methods for the decision algorithm to be called. Finally, in addition to user-perceived metrics, we also study the impact of multiple devices using the QA-MFM approach on the load of the various networks involved.

The remainder of this text is organised as follows. In Section II, we present the mixed-traffic QA-MFM BIP formulation, its utility function and optimisation objective. Section III presents the basic OPNET simulation model, and the multi-user available-capacity estimation methods we use. Mobility models and metrics of interest are also introduced in this section. This simulation model is first used, in Section IV, to calibrate the decision algorithm in terms of scaling factors for the multi-objective utility function and triggering methods. Section V then presents and discusses performance results. Finally, prior to concluding in Section VII, a review of related work is offered in Section VI.

II. Binary Integer Programming-based MFM

A. Multihomed Flow Management Formulation

Given possible access to N network through I interfaces, the multihomed flow management problem consists in finding the distribution of F flows (*i.e.*, over which interface i, connected to network n, each flow f should be sent), and preemptively setting their respective configuration parameters from set C (*e.g.*, video codec or bit-rate) to match the offerred resources, in order to optimise some metrics [3]. An upper bound of the number of possible combination is $F \times C \times I \times N$, though compatibility between interfaces and the available networks, as well as the available parameters of each flows might lead to a smaller actual range. It is therefore possible

978-1-4673-5828-6/13/$31.00 ©2013 IEEE

to create $F \times C \times I \times N$ binary variables of the form [4]

$$x_{fcin} = \begin{cases} 1 & \text{if flow } f \text{ with configuration } c \text{ is distributed on interface } i \text{ connected to network } n, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

to which a few constraints apply,

$$\begin{cases} \forall f, \quad \sum_{c,i,n} x_{fcin} = 1, & \text{(2a)} \\ \forall f,c,i,n, \quad x_{fcin} \leq a_{in}, & \text{(2b)} \\ \forall i, \quad \sum_{n} a_{in} \leq 1, & \text{(2c)} \\ \forall i,n, \quad \left(\sum_{f,c} x_{fcin} C_{fc} \right) \leq C_{in}. & \text{(2d)} \end{cases}$$

(2a) ensures one and only one parameter set and network distribution is active for each flow, while (2b) and (2c) define auxiliary variables a_{in} to constrain each interface to be associated with only one network at a time; (2d) ensures the distribution of flows (where C_{fc} represents the capacity requirement of f when using configuration c) does not exceed the capacity available to the user (see Section III-B).

B. Quality Awareness

Quality awareness is introduced as an objective to optimise specific metrics. High-level user-perceived metrics are closest to the user, and therefore most relevant [3]. They include quality of experience (QoE) Q_{fcin}, as estimated with ITU-T's objective models [5], [6] based on achievable network performance (*e.g.*, capacity C_{in} or delay D_{in}), as well as the battery E_{in} drain and monetary cost M_{in} induced by using network links.

The generic form of the optimisation objective is

$$\forall f, \quad \max \left(\alpha Q_{fcin} - (\beta E_{in} + \gamma M_{in}) \right), \tag{3}$$

where α, β and γ are scaling factors (we investigate how to set their actual values in Section IV-A). More specifically, both energy and monetary costs are assumed to have two components, based on time running (for baseline costs) and the amount of transferred data; the latter are denoted E'_{in} and M'_{in}. For ease of notation, time is factored out of the objective so E_{in} and M_{in} are unit-less, and E'_{in} and M'_{in} are expressed in (Bps)$^{-1}$. The complete formulation of the objective is therefore

$$\max \sum_{f,c,i,n} x_{fcin} \left(\alpha Q_{fcin} - (\beta E'_{in} + \gamma M'_{in}) C_{fc} \right) - \sum_{i,n} a_{in} (\beta E_{in} + \gamma M_{in}), \tag{4}$$

which is not linear due to the presence of products of the binary variables. This can be solved by pre-computing a utility function,

$$\forall f,c,i,n, \quad u_{fcin} = \alpha Q_{fcin} - (\beta E'_{in} + \gamma M'_{in}) C_{fc}. \tag{5}$$

This approach leads to pre-computing all utilities u_{fcin}. In a realistic example case of 5 flows with a total of 10 possible configurations, 2 interfaces and 4 reachable networks, this gives a maximum of 400 values which seems reasonable as the computation only involves basic arithmetic operations. The problem thus becomes

$$\max \sum_{f,c,i,n} x_{fcin} u_{fcin} - \sum_{i,n} a_{in} (\beta E_{in} + \gamma M_{in}), \tag{6}$$

to be optimised under constraints (2a–2d).

C. Mixed Traffic

Formulation (6) works well for application flows which have a known C_{fc} for a given c, such as video or audio streaming ($\vec{f_{rt}}$). It is however not trivial to determine C_{fc} for elastic traffic ($\vec{f_{el}}$, *e.g.*, web browsing or file download), as congestion-controlled transports such as TCP attempt to share the available resource "fairly". The fairness of the capacity distribution depends, amongst others factors, on C_{fc} is therefore dependent on criteria external to the flow f itself, or even the device whose flow f is.

To overcome this issue we introduce a two-step decision mechanism based on the assumption that real-time traffic has a higher priority. It is therefore allocated first. A new utility term q_{in} quantifying the capacity occupancy ratio by the real-time flows on each interface is introduced,

$$q_{in} = \sum_{\vec{f_{rt}}} x_{fcin} \frac{C_{fc}}{C_{in}}. \tag{7}$$

Its objective is to maximise the remaining capacity for the elastic traffic in case of high load (overload) situations.

For the mixed traffic scenarios, (6) thus becomes

$$\max \sum_{f \in \vec{f_{\mathrm{rt}}},c,i,n} x_{fcin} u_{fcin} - \sum_{i,n} a_{in} (\beta E_{in} + \gamma M_{in}) - \sum_{i,n} \delta q_{in}, \tag{8}$$

and the decision process is done as follows.

1) (8) is optimised first, to decide the network associations for real-time flows as well as their parameters and distribution. After this step, the remaining capacity $r_{in} = C_{in} - \sum_{f \in \vec{f_{\mathrm{rt}}}} x_{fcin} \cdot C_{fc}$ is known for each interface. Here the estimation of C_{in} for each network interface is explained in Section III-B.
2) (6) is then optimised for elastic flows, to decide their flow distributions and capacity sharing on each link, taking $C_{in} = r_{in}$, and distributing it fairly amongst the remaining non real-time flows.

III. Simulation Model and Scenario

A. Integration of OPNET and CPLEX

Our simulation scenario follows the proposal from 3GPP specifications [7] in the integration of 3GPP access technology (*i.e.*, LTE) and trusted non-3GPP access technology (*i.e.*, legacy WLAN 802.11g) with host-based mobility solutions (*i.e.*, Dual Stack Mobile IPv6). A simulation model has been implemented using OPNET.[1] As per [7], the home agent (HA) function is located at the Packet Data Network (PDN) gateway. The remote server acts as a correspondent node (CN, see Fig. 1). A comprehensive description of this heterogeneous network simulator can be found in [8]. It should be noted that our focus is only on the downlink access for LTE and WLAN. This implies that no uplink transmissions are performed for WLAN during the simulation. Instead uplink traffic (*e.g.*, TCP ACK packets) is transmitted by the user on the LTE access link.

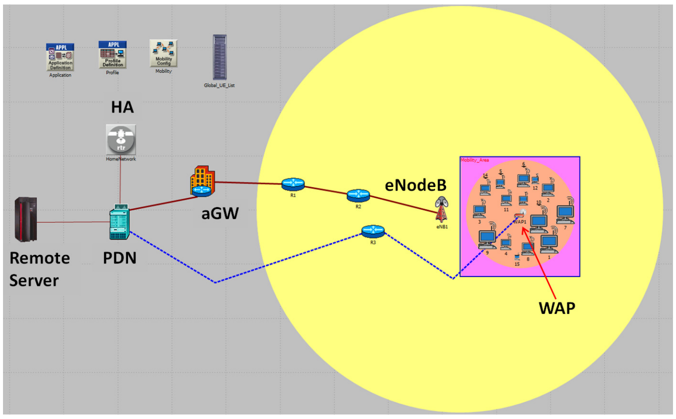

Fig. 1. OPNET simulation scenario. Traffic for N users is distributed at the PDN gateway over both LTE (large, yellow, coverage) and WLAN (smaller, orange, coverage). Users move randomly within the purple square.

The optimisation presented in Section II has been implemented within the CPLEX MIP solver[2] and linked to OPNET using the former's *Callable C library*.

For comparison purposes, we use the most common network selection technique where a mobile device senses the available networks, and favours any WLAN over cellular links, only used as a last resort [2]. We call it 3GPP-HO [8]. In essence, a mobile device will always be connected to the LTE network, but switches to the WLAN link shortly after it becomes available, and keeps using it until it becomes unreachable.

B. Capacity Estimation

The QA-MFM approach relies on estimates of the capacity available to each user on the reacdable networks (C_{in}). To estimate link capacities, most commonly used techniques require data traffic to flow between the user terminal and base station or WLAN AP. Using test data flows for this purpose causes bandwidth overheads, and only provides accurate estimates if data traffic saturates the link. Additionally, information will be missing for user terminals that have just attached to a network and have not yet received any data. Rather, in this work, we use analytical methods to estimate the available link capacity of both WLAN and LTE networks at the PDN. In real implementation, the results of this estimation can be communicated to the mobile users by way of mechanisms such as IEEE 802.21 [9] or OConS [10].

1) C_{in} for the WLAN Link: The PDN manages the WLAN resources for the downlink by performing a quasi-packet-scheduling based on the number of active users as well as their PHY data rates [11]. This scheme resembles the well-known Time Domain Multiple Access, where users are given equal share of time slices during which they can transmit as many packets as their PHY rate allows. Here, we assume that the PDN can obtain the WLAN PHY data rates without delays.

Following this scheduling scheme, the capacity of the WLAN network C_{AP} and that available to a user u (C_{in}) can be computed as

$$C_{\mathrm{AP}} = \frac{\sum_{u=1}^{N} w_u p_u}{\sum_{u=1}^{N} w_u t_u}, \qquad C_{in} = \frac{w_u p_u}{\sum_{i=1}^{N} w_u t_u} \tag{9}$$

where t_u is the time duration required to transmit a packet of size p_u bits to user u operating at a certain PHY data rate and N is the total number of active users in the hotspot and

$$r_u = \frac{p_u}{t_u}, \quad w_u = \frac{r_u}{\sum_{u=1}^{N} r_u}. \tag{10}$$

2) C_{in} for the LTE Link: We determine the radio resources (number of PRBs, Physical Resource Blocks) each user can receive from the LTE scheduler at the eNodeB to estimate the capacity it represents. The LTE eNodeB employs a round robin scheduling in the frequency-domain to schedule the radio resources for all users in the cell. Thus we can assume that each user (one LTE radio bearer per user) will get an equal share of the PRBs. The number of PRBs assigned to one user in a Transmission Time Interval (TTI) is computed as

$$\mathrm{PRB}_u = \frac{\mathrm{PRB}_{\mathrm{total}}}{N_{\mathrm{nGBR}}}, \tag{11}$$

where $\mathrm{PRB}_{\mathrm{total}}$ is the total number of PRBs of an eNodeB cell, and N_{nGBR} is the maximum number of non-guaranteed bit-rate (nGBR) bearers to be served per TTI.

The total capacity C_{LTE} is computed based on the estimated PRB_u and the measured SINR, according to the 3GPP Modulation and Coding Scheme (MCS) table [12]. The link capacity C_{in} for each user is then derived as

$$C_{in} = \frac{C_{\mathrm{LTE}}}{N_{\mathrm{LTE}}/N_{\mathrm{nGBR}}}. \tag{12}$$

where N_{LTE} is the total number of active users in the cell.

C. Simulation Parameters

We consider an LTE network comprising a single eNodeB with a 5 MHz spectrum (25 PRBs), a single cell with a 350 m radius. The Wi-Fi network uses an 802.11g MAC with RTS-CTS enabled and a coverage of 100 m. The wireless channels are modelled with a macroscopic path loss model [13] and correlated slow fading [14], and Jake's like fast fading with user profile ITU-Veh.A. The users adopt a random direction mobility model with 3 km/h movement.

[1] http://www.opnet.com

[2] http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/

D. Demand Model

All users have a similar demand model, composed of 3 web sessions and 2 video flows. The web sessions request objects of a constant size (1 MB) at Poisson-distributed intervals with an inter-arrival time of 30 s. The video flows are sent at 30 fps, with a varying frame-size depending on the codec in use. Four different frame sizes are considered: 1667 B (400 kbps), 2500 B (600 kbps), 3333 B (800 kbps) and 4167 B (1 Mbps); a frame size of 0 denotes that the flow is not sent.

E. Metrics

We evaluate the performance of the approach as well as verify that it scales and does not adversely impact the networks to which the devices connect.

For the user-perceivable quality, we collect the components of the QoS that it receives (namely capacity C_f, end-to-end delay D_f and application packet loss rate L_f). These QoS tuples are then used in the ITU-T's objective quality models for the relevant application flow type [5], [6] to obtain a Mean Opinion Score (MOS) estimate (see [3], [4] for more details). MOS is a scale from 1 to 5, 1 being the worst, 5 the best. Video QoE is measured every second, while web QoE is computed upon completion of each transfer.

The battery life and the final price the user will have to pay are the two other relevant metrics. We reuse the data from [4, table 1.1] for these basic costs.

To evaluate the impact of multiple users employing our proposed decision technique on the reachable networks, we consider usual performance metrics such as the number of active users on a network, the overall capacity usage and load they incur.

IV. Calibration

A. Scaling Factors

There are four scaling factors in (8), α, β, γ and δ, accounting for the difference in range of the terms of the objective. α and δ are the most important as they control the distribution of the capacity and the resulting QoE to the various flows. We therefore attempt to find a partial calibration for our system, keeping $\beta = \gamma = 1$.

We do not report the full results here for the sake of space, but using $(\alpha : \delta) = (1 : 10)$ was found to be the best, providing the best results on both QoEs for elastic and real-time traffic, as well as the monetary cost, while not degrading the battery consumption. We use this parametrisation in the rest of this paper.

Another set of weights is also expected to be provided by the user and/or the access providers with respect to their respective preferences. Though these weights should appear in the same place as scaling factors here, we keep them for future study.

B. Triggering Mechanism

In [4], the solving time for one pass of the decision algorithm was well below a second, even for rather large problems. Here, we investigate what periodicity for subsequent calls to the optimisation works best. We also consider an aperiodic *on-demand* triggering mechanisms, where a redistribution of the flows is done whenever an application flow arrives or finishes , as well as when the available network capacity changes by certain ratio (we use 10% of the current capacity in this work).

Fig. 2 shows multiple metrics related to the QoE of application flows for the periodic triggering method with various periods as well as for the on-demand scheme. Similarly, Fig. 3 shows the impact on the battery consumption and total price, as well as the number of calls to the solvers for each case.

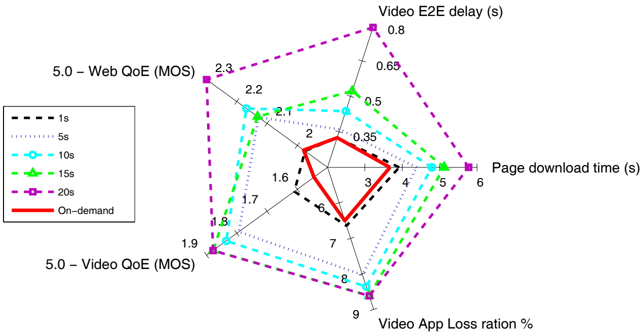

Fig. 2. Influence of the triggering mechanisms and their period on QoE-related metrics for both types of flows. The on-demand method performs best.

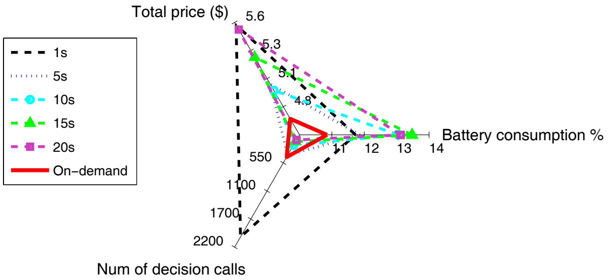

Fig. 3. Influence of the triggering mechanisms on costs and the number of decision calls. Once again, the on-demand method provides the best overall performance.

It is clear that the on-demand scheme out-performs the periodical triggering mechanism with any period range, consistently achieving higher-than-3 average QoEs for both flow types, while limiting both costs and reducing the number of needed calls to the solver. We therefore select this method and use it for the rest of this paper.

V. Results and Discussion

A. Statistical Results

We now compare the QA-MFM proposal to the more classical 3GPP-HO. We vary the number of users from 1 to 15, and measure the performance both network selection techniques achieve with respect to the metrics we defined in Section III-E. Results are collected and averaged over 5 simulation runs of 2000 s (resulting in 10,000 samples for per-second metrics, or $5 \times F$ samples for per-flow metrics). Error bars show the 95% confidence interval for the mean.

1) Video Flows: Fig. 4 shows the average QoE, measured every second, for video flows. The QA-MFM approach consistently achieves a significantly higher QoE than the 3GPP-HO.

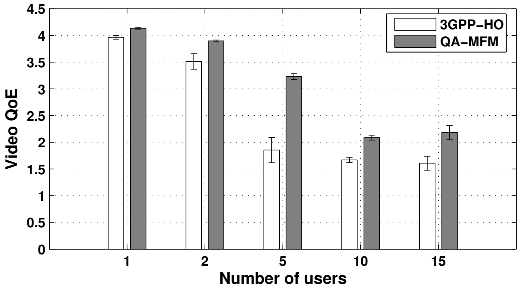

Fig. 4. Comparison of the quality of experience for video flows achieved by both techniques.

Investigating the application parameters, in Fig. 5, we see that the QA-MFM technique achieves this improvement over 3GPP-HO by preemptively setting the bit-rate of the video flows to what the new access network supports given its current load. Fig. 6 shows that this also greatly reduces the loss rate for application data.

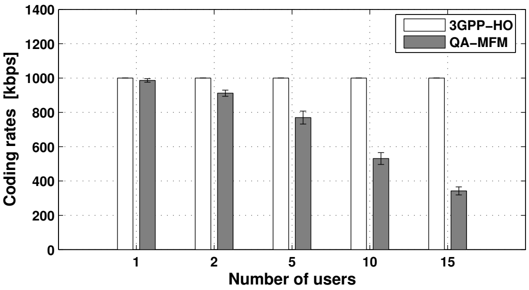

Fig. 5. Codec rates in use. The 3GPP-HO technique is oblivious of this parameter, while the QA-MFM approach tries to match them to best match the network conditions.

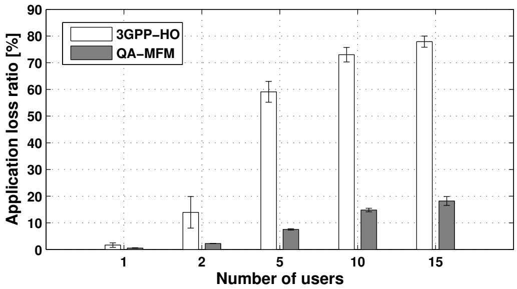

Fig. 6. Application loss rate for the video flows. With lower coding rates, the QA-MFM approach succeeds in reducing the loss rates on more congested networks.

Though this diminishes the QoE, reducing the bit-rate of video flows in order to reduce the loss rate, as the QA-MFM approach does, provides a better trade-off than trying to send the highest video quality at all times. As shown later in section V-A4 this also allows to not overload or create an imbalance between the access networks. The remaining losses shown in Fig. 6 are due to inevitable collisions in congested wireless networks.

2) Web Traffic: Fig. 7 shows the perceived QoE for the web sessions once they terminated, using a 15 s expected completion time in the formulas from [5], out of all flows in 5 simulations runs.

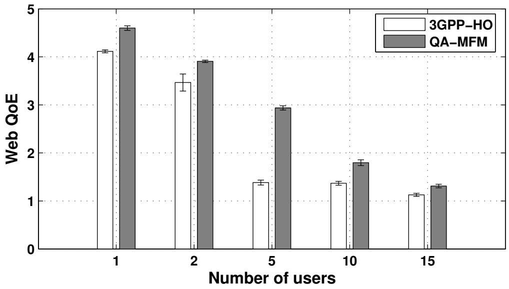

Fig. 7. Comparison of the quality of experience for web sessions achieved by both techniques.

From a qualitative standpoint, the QA-MFM approach again achieves a significantly higher QoE, by maintaining download times lower (as seen in Fig. 8).

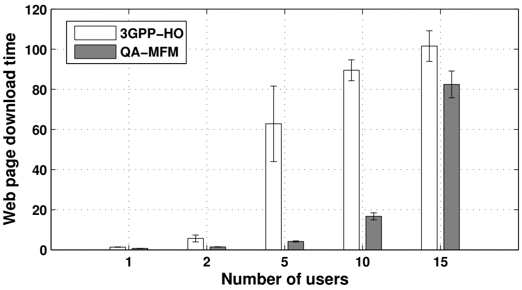

Fig. 8. Download times for web sessions achieves by both approaches. As the QA-MFM technique can choose to use more than one network at once, it can more effectively distribute the current traffic, and achieve lower transmission times.

3) Other Costs: Fig. 9 and 10 show the battery consumption and the price incurred by each techniques. These metrics are collected out of 5 simulation runs.

Unlike the previous metrics, QA-MFM does not perform as well here, and the 3GPP-HO technique achieves lower costs. We however recall from section IV-A that the scaling factors for these two metrics were not calibrated.

4) Network Load: We now consider the impact of both techniques on the wireless networks in use. Fig. 11 shows the number of users using each network for both approaches, as measured once every second. The sum of the metrics for QA-MFM might exceed the actual number of users in the scenario as each user can use more than one network at the same time.

The 3GPP-HO scheme always prefers the WLAN network when available, with most users only using this one, while the

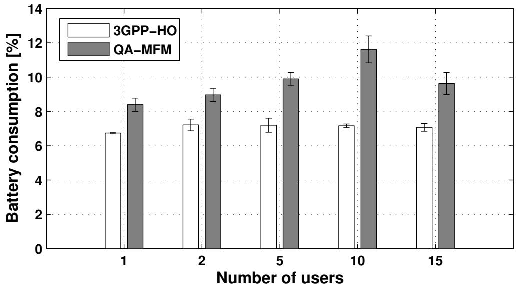

Fig. 9. Average battery consumption at the end of a 2000 s run. 3GPP-HO achieves better results here, conserving more energy.

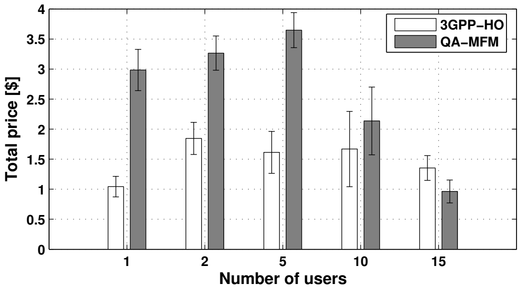

Fig. 10. Average price incurred at the end of a run. Again, 3GPP-HO performs better, saving more money for the user.

QA-MFM achieves a more balanced distribution, as seen in Fig. 12. Moreover, with an increasing number of users, the loads do not explode as is the case with 3GPP-HO, owing to the proper adjustment of applications parameters.

B. Discussion

The results just presented show that the QA-MFM managed to achieve a better application quality for both types of flows. However, it was not as efficient with respect to the costs, where the 3GPP-HO maintained lower energy consumption and price.

This can be explained by the fact that, while the 3GPP-HO technique only uses one access network at a time, the QA-MFM technique has the opportunity of using multiple links at once. This is confirmed in Fig. 11, where QA-MFM users connect to more than one network at once. 3GPP-HO is therefore at a clear advantage here, due to its limitation of choosing only one network for all its traffic.

It would be interesting to make similar comparisons with techniques other than 3GPP-HO which also consider the simultaneous use of more than one network. Also, we recall from section IV-A that we did not attempt to calibrate scaling factors β and γ, related to the costs. Both tasks are kept for future work.

Nevertheless, as it takes application parameters into account, and matches them to the available network capacity, the QA-MFM approach manages to reduce the application losses, as well as the load on the network it uses. It also balances users and their traffic more fairly across the available networks.

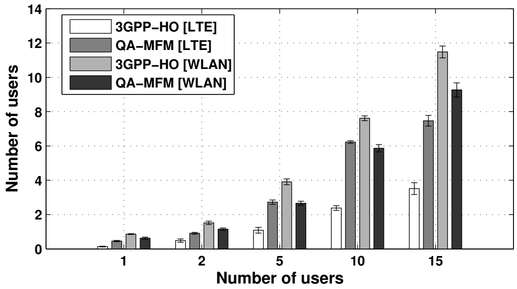

Fig. 11. Repartition of users on both LTE and WLAN depending on the technique in use. For QA-MFM, the sum is higher than the total number of users as they can use more than one network at once.

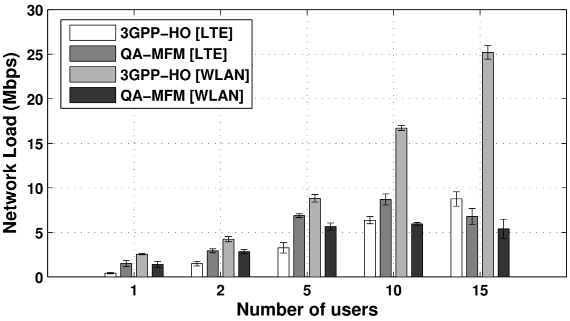

Fig. 12. Loads generated on the LTE and WLAN depending on the approach. While the 3GPP-HO load increases with the number of users, the QA-MFM technique manages to maintain it at lower, more stable, values.

VI. Related Work

This section reviews work related to metrics used for network selection as well as various decision techniques. A more detailed review can be found in [15, chap. 2, sec. 2.3].

a) Criteria for Network Selection: A large range of criteria has been proposed to discriminate access links and networks in order to select the best ones to connect to. The simplest mechanisms are based on measuring the quality of the radio signal (*e.g.*, signal-to-noise ratio or received signal strength) [16], delay or data rate [17], [18]. and comparing them to a threshold. More relevant than the link layer properties for communication facilitated by transport protocols like TCP, end-to-end parameters such as network path capacities or RTTs are important to support feature-rich applications [19], [20], [18], [21], [22], [23], [24]; some proposals also specifically take the application requirements into account in this phase [17]. Additionally, the reachability of the Internet [20] has also been proposed as a criteria in this case.

As we argue in this paper, battery life is important in a mobile context, and trade-offs have been considered to preserve it [21], [24], [23]. Similarly, multiple approaches take monetary considerations into account [19], [17], [21], [23], [25]. The currently observed application layer performance can also be used as an indication of the "health" of a network link [25]. However, QoE is still very rarely used for such tasks.

b) Flow Distribution: Two main classes of solutions can be distinguished. The first group applies traffic classification and load balancing approaches of conventional wired technologies after network uplinks have been selected and established. Simple policies, based on flows' destinations or port, to decide which network is the most appropriate are often seen [26]. However, more complex techniques proposed distribute new flows with more elaborate heuristics [18]. Approaches in the second class take a more holistic approach by performing network selection and flow distribution at the same time. A number of solutions rely on knowledge of the applications' requirements to select the network which most closely matches them [19], [23]. These approaches however come at the cost of a larger solution space to search. To address this issue, the problem was modelled as a Markov chain [27] to leverage decision process techniques of that field. Binary integer programming techniques have also been proposed [21].

VII. Conclusion and Future Work

In this work, we refined an earlier proposal of an application quality-aware network selection and flow distribution system. We extended prior formulations into a workable model encompassing both real-time and elastic traffic, and introduced more accurate capacity-estimation techniques for LTE. We provided a first attempt at calibration of the main parameters of our linear programming objective, and its integration into a real system in terms of how often the solver should be called.

We showed that our technique, being aware of application parameters, is able to more accurately adapt them to the current network conditions. However, the use of multiple networks causes higher costs which the 3GPP-HO techniques did not incur. We also found that the QA-MFM technique distributes application flows in a more balanced way over the available networks.

Future work will investigate full calibration of the proposed technique, as well a further refinements. We will also compare our approach to other multi-homing techniques in order to get better insight about the respective cost uses. Finally, we want to conduct larger scale and more realistic simulation scenarios, with a larger number of users and networks.

Acknowledgements

This research has received funding from the European Union Seventh Framework Programme (FP7) under grant agreement n°257448, *Scalable and Adaptive Internet soLutions (SAIL)*. Ramón Agüero would also like to express his gratitude to the Spanish Government for its funding in the project C3SEM (TEC2009-14598-C02-01).

References

[1] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Communications*, vol. 10, no. 1, 2003.

[2] M. Wasserman and P. Seite, "Current practices for multiple-interface hosts," Internet Requests for Comment, RFC Editor, RFC 6419, 2011.

[3] O. Mehani, R. Boreli, M. Maher, and T. Ernst, "User- and application-centric multihomed flow management," in *LCN 2011*, 2011.

[4] X. Li, O. Mehani, R. Agüero, R. Boreli, Y. Zaki, and U. Toseef, "Evaluating user-centric multihomed flow management for mobile devices in simulated heterogeneous networks," in *MONAMI 2012*, vol. 58, 2013.

[5] "Estimating end-to-end performance in IP networks for data applications," ITU-T SG12, Recommendation G.1030, 2006.

[6] "Opinion model for video-telephony applications," ITU-T SG12, Recommendation G.1070, 2007.

[7] "Architecture enhancements for non-3GPP accesses," Release 11.2.0; Also published as ETSI TS 123 402, TS 23.402, 2012.

[8] U. Toseef, Y. Zaki, A. Timm-Giel, and C. Görg, "Uplink QoS aware multi-homing in integrated 3GPP and non-3GPP future networks," in *MONAMI 2012*, 2012.

[9] E. Piri and K. Pentikousis, "IEEE 802.21," *The Internet Protocol Journal*, vol. 12, no. 2, 2009.

[10] R. Agüero, L. Caeiro, L. M. Correia, L. S. Ferreira, M. García-Arranz, L. Suciu, and A. Timm-Giel, "OConS: Towards open connectivity services in the future Internet," in *MONAMI 2011*, 2011.

[11] U. Toseef, Y. Zaki, L. Zhao, A. Timm-Giel, and C. Görg, "QoS aware multi-homing in integrated 3GPP and non-3GPP future networks," in *ICSNC 2012*, 2012.

[12] "Physical layer procedures," Release 9.3.0, 3GPP/TSG R WG1, TS 36.213, 2012.

[13] "Physical layer aspect for evolved universal terrestrial radio access (UTRA)," Release 7.1.0, 3GPP/TSG R WG1, TR 25.814, 2006.

[14] Y. Zaki, T. Weerawardane, C. Görg, and A. Timm-Giel, "Multi-QoS-aware fair scheduling for LTE," in *VTC 2011-Spring*, 2011.

[15] O. Mehani, "Contributions to mechanisms for adaptive use of mobile network resources," Ph.D. dissertation, Mines ParisTech / University of New South Wales, 2011.

[16] S. Mohanty and I. F. Akyildiz, "A cross-layer (layer 2 + 3) handoff management protocol for next-generation wireless systems," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, 2006.

[17] M. Alkhawlani and A. Ayesh, "Access network selection based on fuzzy logic and genetic algorithms," *Advances in Artificial Intelligence*, vol. 2008, 2008.

[18] S. Kandula, K. C. Lin, T. Badirkhanli, and D. Katabi, "FatVAP: Aggregating AP backhaul capacity to maximize throughput," in *NSDI 2008*, 2008.

[19] V. Gazis, N. Alonistioti, and L. Merakos, "Toward a generic "always best connected" capability in integrated WLAN/UMTS cellular mobile networks (and beyond)," *IEEE Wireless Communications*, vol. 12, no. 3, 2005.

[20] A. J. Nicholson, Y. Chawathe, M. Y. Chen, B. D. Noble, and D. Wetherall, "Improved access point selection," in *MobiSys 2006*, 2006.

[21] V. E. Zafeiris and E. A. Giakoumakis, "Mobile agents for flow scheduling support in multihomed mobile hosts," in *IWCMC 2008*, 2008.

[22] J. Pang, B. Greenstein, M. Kaminsky, D. McCoy, and S. Seshan, "Wifi-Reports: Improving wireless network selection with collaboration," in *MobiSys 2009*, 2009.

[23] J.-M. Bonnin, "La diversité technologique au service des terminaux et routeurs multiconnectés," Mémoire d'habilitation à diriger des recherches, Institut Télécom—Télécom Bretagne, 2008.

[24] B. Xing and N. Venkatasubramanian, "Multi-constraint dynamic access selection in always best connected networks," in *MobiQuitous 2005*, 2005.

[25] K. Piamrat, C. Viho, A. Ksentini, and J.-M. Bonnin, "QoE-aware network selection in wireless heterogeneous networks," Inria, Tech. Rep. RR-7282, 2010.

[26] M. Tsukada, O. Mehani, and T. Ernst, "Simultaneous usage of NEMO and MANET for vehicular communication," in *TridentCom 2008*, 2008.

[27] J. Singh, T. Alpcan, P. Agrawal, and V. Sharma, "A Markov decision process based flow assignment framework for heterogeneous network access," *Wireless Networks*, vol. 16, no. 2, 2010.