Differentiated Classes of Service and Flow Management using An Hybrid Broker

João Silva, Rui Neto Marinheiro, José Moura and João Almeida

Instituto Universitário de Lisboa (ISCTE-IUL), Instituto de Telecomunicações, Portugal
{joao.silva, jose.moura, rui.marinheiro, jmdaa}@iscte.pt

Abstract— Recently, mobile networks have been overloaded with a considerable amount of data traffic. The current paper proposes a management service for mobile environments, using policies and quality metrics, which ensure a better usage of network resources with a more fine-grained management based on flows with different classes of service and transmission rates. This management of flows is supported through a closed innovative control loop among a flexible brokerage service in the network, and agents at the mobile terminals. It also allows the terminals to make well-informed decisions about their connections to enhance the number of connected flows per technology and the individual service level offered to each flow. Our results indicate that the proposed solution optimizes the usage of available 4G network resources among a high number of differentiated flows in several scenarios where access technologies are extremely overloaded whilst protecting, through a low complexity scheme, the flows associated to users that have celebrated more expensive contracts with their network operators.

Index Terms—Flow management, Brokerage service, Offloading, Load balancing, Heterogeneous network, Differentiated management policies.

I. INTRODUCTION

According to the Cisco Global Forecast [1], in 2012, more than 50 percent of the total mobile traffic was video, average smartphone usage grew 81 percent, and globally 33 percent of total mobile data traffic was offloaded onto the fixed network through Wifi or femtocells. This traffic offloading occurs due to the lack of capacity in the mobile network infrastructure, originally dimensioned to support only voice and messages. In our opinion a realistic strategy to avoid congestion in mobile networking is establishing national roaming contracts among operators, enabling customers from an operator to use other operators’ networks, when the former operator’s network lacks resources to satisfy the user demand. The deployment of femtocells can also enhance the network coverage of a wireless technology owned by a single mobile operator.

A very recent work [15] proposes that a mobile multimode terminal can use all the available connectivity options. The mobile terminal can choose dynamically the more suitable network to obtain a seamless wireless connection, obtain faster connections by stitching flows over multiple networks, decrease the usage cost by choosing the most cost-effective network that meets application requisites, and reduce the energy consumption by selecting the technology with the lowest energy-usage per byte.

The current paper follows on previous work [2-5]. It is currently suggested a brokerage system that manages the network architecture, in which distributed nodes discover relevant context information to enhance the usage of local available connectivity resources [6]. We also advocate that multi-interface handheld terminals will soon have the battery autonomy and the capability to perform network access using simultaneous multi-radio access technologies (RAT). In addition, of particular interest is the support of simultaneous data/multimedia flows through different access systems (LTE, WLAN, Wimax). A flow is a sequence of packets between end-host applications that transport data in the same (virtual) session. Flows have different characteristics and/or network requirements, and adding to this, the simultaneous access capability of terminals, it is possible to better manage the available network resources by providing policies to terminal IP flow routing via different access systems; enabling the move of IP flows between access systems (offloading) in a seamless way (with a few errors and low latency). In addition, it is possible to balance the load among available technologies.

The first new aspect of our work considers the management of flows, where each flow has an individual network access selection process, instead of a single selection process for the user/terminal [5]. Secondly, the current work introduces a novel functionality based on the upgrade/downgrade (UD) of the rate of some selected flows, depending on the available network resources and load. Thirdly, this work proposes two distinct flow types, designated by gold and normal, each one with different flexibility on the fulfillment of their contractual requirements. In fact, the requirements of gold flows should be always fulfilled except with the particular case when there are no available networks resources and all the connected flows are gold. Alternatively, the requirements of normal flows are only fulfilled if there are available network resources and no gold flow is negatively impacted on. In this way, our work has the pertinent advantage of supporting the contractual quality of service requirements of the gold flows in spite of network congestion situations, using a low-complexity distributed management solution. This functionality is very important for network operators because the gold flows are associated to users with more expensive contracts. Finally, the channel probabilistic model for dropping flows is enhanced with a new decision level based on the available network resources. Combining all these new functionalities, our proposal dynamically balances flows.
among technologies [7] in a distributed way, and uses more efficiently the available resources from the heterogeneous network infrastructure among a high number of flows belonging to distinct user types.

The section II contextualizes and highlights the novel aspects of our work in the literature. In section III is discussed the scenario and requirements of the current management proposal. The section IV describes our solution. In section V it is evaluated our proposal, discussing its more relevant achievements. Finally, section VI summarizes our research contributions and points out some guidelines to future work.

II. RELATED WORK

The current work discusses a brokerage system that manages congestion in a network access environment, balancing high data demand among distinct wireless access technologies, and guaranteeing a connection quality to mobile flows, depending on their contractual requisites. The connection quality is measured as the wireless bitrate used by each flow. The flow mobility is mainly due to balance dynamically the traffic load among the access technologies deployed in a specific location.

However, the first approach to the problem has been to perform an inter-technology handover between available technologies, with all the traffic routed through one or the other access as whole. A survey about mobility is available in [8], and network brokers in [5] that follow on this idea. But a more interesting scenario is to have the capability to move selected IP traffic (e.g., Web traffic, VoIP) while supporting simultaneous access seems a more attractive solution. In this way, mobile operators can develop policies for IP flow mobility, and control which traffic is routed over different access technologies.

This paper contributed to the perspective of integrating complementary access technologies with overlapping coverage to provide the expected ubiquitous coverage and to achieve the Always Best Connected (ABC) concept [9]. This concept allows a flow to use at any time the more suitable access network. This work is a tentative to propose an ABC proposal, satisfying the following functional requisites: load balancing, flow management among networks, flow admission, flow contractual types (e.g. normal and gold flows) and resource sharing. In addition, this work supports at the network the dynamic calculation of Network Attachment Point (NAP) quality in terms of the wireless status. Further, the terminals support a dynamic cost function to rank target NAPs based on the connection quality offered by each one of these. In this way, the final decision in terms of NAP selection per flow is issued at terminals and supported by the distributed broker units at the network infrastructure.

The current broker architecture has an innovative hybrid design, aggregating both centralized and distributed functionalities in comparison with [10]. In addition, the current proposal is more flexible than [10] because the former can operate in both inter-domain and intra-domain networking scenarios. This proposal also extends [11]. In fact, this work is able to control a network infrastructure, composed by any access technology, with the novelty of downgrade/upgrade of flows according the available resources and load demand, enhancing flow network selection and admission.

Our proposal can be easily integrated with IEEE 802.21 [16], which it is a very promising deployment architecture for future access networks, supporting different functionalities such as signaling protocol, NAP discovery, network selection, security and QoS/QoE in heterogeneous environments. Nevertheless, as stated above, almost all previous approaches have only considered mobility with handovers of all flows from one technology to the other as a whole, implemented either using host-based mobility management such as Dual Stack Mobile IP [20] or network-based localized mobility management such as Proxy Mobile IPv6 [21]. Only recently alternative solutions [12, 15, 18, 19] are trying to overcome this limitation and Internet engineering task force is working at the present on Proxy Mobile IPv6 Extensions to Support Flow Mobility - [17]. The results obtained with the research presented here can be used to take flow handover decisions implemented by these extensions to Proxy Mobile IPv6.

III. SCENARIO AND REQUIREMENTS

This section discusses the scenario, the addressed problems and the functional requirements for the heterogeneous network infrastructure. A single location is covered by several wireless access networks, each one administered by a distinct provider. The load may vary over time in terms of the number of flows requiring connection, demand characteristics (e.g. audio/video, high/low bitrate), and available resources of each access network may not guarantee a good connection quality for a high number of flows. We assume a cooperative model among network providers, coordinated by a brokerage service on the edge of the network, to enhance the service of wireless connectivity provided by a heterogeneous network infrastructure. This work attempts to balance an extremely high number of multimedia flows across available access networks, applying management policies, and offering a satisfactory connection service to the maximum number of accepted flows.

Fig. 1 shows several brokers deployed at convenient nodes of the infrastructure. The brokerage functionality is distributed amongst Points of Presence (PoPs) and access networks. A PoP is a location where several wireless access technologies exchange their traffic with the Internet through backhaul links. The proposed brokerage service manages the edge of the network infrastructure on behalf of the network providers. Any broker unit can coordinate its decisions with other remote brokers located at neighbor PoPs in a peer-to-peer communication (link 6 of Fig. 1).

We assume that each customer obtains a connection, paying a fee, dependent on the subscription contract with his provider. In addition, the roaming cost is shared (or paid) among the providers, transparently to the roaming users. Second, the brokerage service evaluates the connection quality associated to each NAP. Third, the brokerage service de
tects any congested link. Fourth, the broker remediates a congestion situation applying adequate policies to the network infrastructure, using a reactive mode. Finally, another predictive mode can be considered for the brokerage service. This mode can evaluate the success/failure of each policy, produce a ranked list of policies, learn the causes and solution for any problem, and predict the next problem.

The broker units are simple software routines running at already deployed network units like Wifi Access Router and Wimax ASNGateWay. In addition, 802.21 can be used to enable the brokerage service to manage several access technologies, either locally at a single network node or remotely between distinct network nodes [6].

IV. Solution Design

The design of our brokerage service to enhance the individual service level offered to each flow is now discussed. Fig. 2 shows the decision thresholds (policies) of the brokerage service to control a wireless access technology: admission of new flows, management of accepted flows accordingly the network status and flows dropped by the physical channel. Each technology has a maximum value of connection rate designated by Max_rate.

Fig. 3 shows the brokerage service modeled as a distributed algorithm that runs continuously in consecutive iterations. Due to the simulation of rush hour traffic, with an excessive amount of flows logging simultaneously, an iterative model was chosen, where a single iteration deals with a specific number of new flows, blocks, drops and flow terminations. The following functional aspects will be discussed: network selection and flow admission, drop of flows and management of connected flows.

There are two flow types, gold and normal, each one with distinct contractual requirements. Consequently, the brokerage service manages distinctively these two flow types. In this way, the brokerage service decision thresholds of Fig. 2 related with flow admission and flow management are only valid for normal users. By other words, gold users are always admitted, if there is available capacity for that, and the rate of a gold user is never changed by the brokerage service.

From Fig. 3 it is visible that there are two distinct reasons to drop flows. The first one is controlled by the brokerage service (Fig. 4). In this way, the brokerage service uses an algorithm that operates in several steps as follows.

First, the brokerage service selects all the normal flows. Secondly, it selects the normal flows with high data rates. Thirdly, it restricts the selection to the most recent flows. Fourthly, the brokerage service drops the normal flows that have been selected through the previous steps. After, if the load is still too high then the four previous steps are repeated again but now for the gold flows.

The second reason why flows can be dropped is due to random channel problems (e.g. interference). This problematic is explained below together with Fig. 7.

The flowchart of Fig. 5 is useful to discuss network selection and flow admission. The load of each technology is available through the entire network, with no delay. The algorithm can also block some flows when the network load is higher than the value Overload. If the broker is disabled, the terminal randomly selects an access network. This assumption was made because depending on the terminal location any of the available access technologies can be selected, potentially the one offering the best Signal to Noise Ratio (SNR). Alternatively, when the broker is enabled, both network selection and admission control are enhanced with using policies with higher preference to a specific technology in detriment of others, e.g. “T2/T1 factor” in Fig. 5. In the case T2/T1=2, the technology T2 is selected when both technologies have a similar load because T2 has normally a better quality coverage than T1. In addition, the brokerage service allows the technology T2 to accept more load than T1 because the threshold associated to the flow admission functionality depends on T2/T1, as illustrated in (1). Thus being, the total load of T1 when T2/T1=2 is half of T2, and the total system capacity

\[ \text{Overload} = 0.8 \times \frac{\text{Max Rate}}{\text{T2/T1}} \]

Fig. 6 describes with more detail the rules to accommodate a set of new flows using a best-effort mode through several steps. Initially, the brokerage service selects the wireless access technology with more available capacity to offer a network connection. Secondly, the brokerage service accommodates all the gold flows in the chosen technology even if the overload threshold is exceeded. This situation could
lead, later on, during the current algorithm’s iteration, the brokerage service to downgrade some connected normal flows to normalize the load. This behavior is acceptable because normal flows are normally associated to cheaper contracts than gold flows are associated to. Thirdly, the brokerage service accommodates all the new normal flows while the instantaneous technology load is below the overload threshold. Otherwise, the brokerage service blocks the remaining normal flows.

We discuss now in what situations the flows (from any type) are randomly dropped, as shown in Fig. 7. A flow drop can occur due to problems (e.g. interference) in a very busy communications channel where some flow connections through that channel are abruptly interrupted. The current work considers three distinct probabilities for dropping a flow, enhancing previous work [5]. Each drop probability is associated to a distinct level of network load (Fig. 2). It is also assumed, for the same load level, a distinct drop probability per technology, depending on the priority parameter T2/T1. The drop probability depends on the load of each technology. In this way, the channel drops flows per technology with a probability of 0.01% if the load of that technology is below the Quality Threshold (QT) associated to it. Otherwise, if the load is inside the range [QT, Overload] then, the channel drops flows with a probability of 0.02%. Alternatively, the channel drops more flows with a probability of 0.5% because the wireless access technology is excessively overloaded. These channel drop values were estimated from
OFDM Bit Error Rates applied to LTE-A [13] [14].

The rate management of connected normal flows is managed with policies that depend on the channel overload to reduce the number of blocked flows, as shown in Fig. 8-9. In respect to the gold flows, the rate of these flows is never changed by the brokerage service because these are protected in the case of network congestion.

Fig. 8 visualizes the flowchart to downgrade a normal
multimedia flow. Initially, the channel overload is evaluated. Then, the flows to be downgraded are selected as few as possible, starting from the ones associated to the service with the highest rate, and the oldest (or newest) flows, in terms of the connection starting time, until the overload ends. The flow downgrade means the connection rate is slightly decreased, offering additional bandwidth to other flows. This increases QoE of other connected flows and allows the acceptance of additional flows, which otherwise will be blocked.

Fig. 9 shows how normal flows can increase the connection rate if there is available network capacity. This scenario occurs when some flows stopped their connections, freeing some bandwidth, which can be reassigned to some flows, enhancing their QoE. In this way, the first flows upgraded are the ones previously downgraded and associated to the service with the lowest rate, and the oldest (or newest) flows.

V. Evaluation

This section discusses the more relevant obtained results from the brokerage service evaluation when the wireless access technologies are severely overloaded. The heterogeneous wireless access scenario of Fig. 1 was used, with two technologies: T1 (LTE-A) and T2 (WLAN-Wifi). Both of these offer a best-effort service in a differentiated way to each one of the two possible multimedia flows (i.e. normal and gold) via multimode terminals. This evaluation scenario is currently very popular in the mobile networking arena, overwhelmed with multimedia data traffic, essentially video. The maximum connection capacity of each wireless technology is 150 Mbps, assuming that both support MIMO. The average rate of flows requiring a network connection is 10 new flow/s (Poisson probabilistic distribution, variance of 5 flows).

<table>
<thead>
<tr>
<th>Table I. Flow Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario</td>
</tr>
<tr>
<td>Rate (Kbps)</td>
</tr>
<tr>
<td>Average duration (time)</td>
</tr>
<tr>
<td>Average flow occurrence (%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II. Evaluation Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

\[
\frac{T1 + T2}{2} \quad (2) \quad \frac{T1 + T2}{3} \cdot 2 \quad (3)
\]
A flow is one of four possible flow types, visualized in Table 1. As an example, in each simulated scenario, 50% of the total number of flows is from type Audio-L, i.e. Audio – Low rate. In addition, Audio-L flows have a rate of 12.2 Kbps and 80s of average time duration. The duration of a flow follows a Log-normal distribution with a variance of 5s. A MATLAB model was built to evaluate the six scenarios listed in Table 2. All the discussion here is driven by average values from 1000 simulation results per scenario. All the results are normalized following (2) or (3), respectively for the scenarios where $T2/T1=1$ or $T2/T1=2$. One should notice that for the scenarios where $T2/T1=2$ the maximum connection capacity of $T1$ is reduced to half value, which is 75 Mbps, and the corresponding value of $T2$ stays unchanged in 150 Mbps.

Results address the next pertinent functional aspects: total network load, number of flows, blocked flows, channel flow drops, and number of downgraded/upgraded normal flows. Further, the results of the current section are discussed using diverse trends (Fig. 10-14) during a total simulation of 1500s. The simulation time is composed of 2 phases; a loading phase of 10 flows/s for the initial 1000s; and an unloading phase absent of flows from 1000-1500s. From Fig. 10, the loading phase is visible for the initial 200s, when the new load is intelligently balanced among the available access technologies. From 200s to 1000s, both technologies become significantly overloaded and an increased number of new flows is either downgraded (case of normal flows), eventually blocked by the system or suffer a channel drop by the influence of an extremely busy communications channel.

These negative actions counterbalance the growth rate of connected flows, which justifies the flat trend, between 200s and 1000s, with equilibrium between flows entering and leaving from each technology. After 1000s, the flows leaving both access technologies are not replaced by new flows and
the network load naturally decreases. In addition, it is good to remind that in the $T2/T1=2$ case, the total load ($T1+T2$) is only $\frac{3}{4}$ of the $T2/T1=1$ case; and thus the normalization of $\frac{2}{3}$ (results are shown for the normalized sum of $T2+T1$).

Fig. 10 shows that after 500s all the available connection capacity is entirely used by the users in the scenarios $G0\%$ (only normal flows) and the gold users in $G50\%$ (50% of new flows are gold). In fact, these two scenarios show that the brokerage service always tries to maximize the use of connectivity capacity in a prioritized way amongst the flows. In addition, the quality of service of gold flows is much more protected than the one of normal flows. To demonstrate this, and observing Fig. 10, the scenario $G10\%$ has only a few new gold flows but these flows practically receive the same amount of connectivity resources than the one received by all the normal flows of scenario $G50\%$. Note that the normal flows of $G50\%$ are almost all downgraded to a minimum value. This occurs in this way because the gold flows should operate at maximum data rate.

Looking at Fig. 11, several interesting things are possible to observe. For the $G0\%$ case, only normal users are present. The fact that there are more flows for $T2/T1=2$ is that the results are normalized and that there are more flows with less bandwidth than bigger flows. This is due to the system, when operating at full capacity, only allowing new flows to be small ones. For the $G50\%$ case, it can be seen that the number of normal and gold flows are almost the same; the difference is that the gold flows weren’t downgraded, and practically all of the normal flows were downgraded (see Fig. 14).

Analyzing now the results visualized in Fig. 12, and more
particularly the scenario G50%, one can conclude that only normal flows are blocked by the brokerage service. The number of blocked flows is marginal, which means that the broker is able to accommodate almost all flows, and rarely operates over the overload marker (the rate of channel drops above the overload marker also “helps” in keeping this balance).

As we already explained before the brokerage service tends to accept all the new gold flows up to the maximum rate of each technology, following a management policy that aims to honor their contractual requirements because the gold users normally pay more money than the normal users. However, since the policy is to let gold users in until the maximum capacity of the system is reached, some gold (alongside normal) users might be dropped due to bad channel conditions (Fig. 13). Remember though that this only happens in cases of a major overload condition; and that all normal users only get in the system after all gold have entered, and if the system is operating under overload. The fact that the channel drops are somewhat marginal warrants this behavior (scenario G50%).

As previously referred, from Fig. 14, it is perfectly visible the downgrade of the rate of normal flows (scenario G50%), giving way to the coming of the new Gold users each iteration (the gold flows are neverdowngraded by the brokerage service). The upgrades occur only after all admissions are off (after instant 1000). Using the G0% as a baseline, notice that the G10% itself already has a big effect on downgrades (especially the case of T2/T1=2, due to a smaller overall system capacity).

VI. CONCLUSIONS AND FUTURE WORK

The research presented in the current paper significantly innovates on the management of multimode terminal accesses at the granularity of flows, by means of a distributed brokerage service in a heterogeneous wireless access infrastructure overwhelmed by multimedia data traffic. In addition, the brokerage service applies distinct management policies to flows according the existing contractual requirements among users and mobile operators.

Our results show that, using this novel management approach, the network performance is improved in the case the connectivity resources are scarce, as follows: the load of flow is intelligently and efficiently balanced among the available access technologies, the number of blocked flows is notably marginal and the gold users have their contractual requirements protected by the brokerage service in detriment of a controlled degradation on the connection quality offered to some normal flows.

Future works can address scenarios where flow management ensures energy efficiency at the terminal and flow mobility is enhanced with contextual information [6].

REFERENCES


