MOTA: Engineering an Operator Agnostic Mobile Service

Supratim Deb
Alcatel-Lucent USA
supratim.deb@alcatel-lucent.com

Kanthi Nagaraj [*]
University of California, Los Angeles
kanthi@ucla.edu

Vikram Srinivasan
Alcatel-Lucent Bell Labs India
vikram.srinivasan@alcatel-lucent.com

ABSTRACT

There are two emerging trends in the mobile data world. First, mobile data is exploding at a rapid rate with analysts predicting $25 - 50\times$ growth by the year 2015. The second trend is that users are demanding greater degree of flexibility in selecting their operators at fine timescales. Across Asia, dual-SIM phones have become popular, while Apple is rumored to be designing a *Universal SIM* that will allow iPhone users to toggle between different operators. This latter trend points towards an impending disruption in wireless service models which could also be the need of the hour from the spectrum shortage perspective.

This points towards a new service model where users can choose an operator based on application needs. However, if users make this choice greedily without network assistance, it can exacerbate spectrum scarcity and degrade user experience. In this work, we consider user devices with multiple network interfaces (3G, LTE etc.) that can be simultaneously active and each running multiple applications. We propose the MOTA service model to enable users to associate each interface with the operator of choice at fine time scales. Under the MOTA service model, through concise signalling information, operators provide information about their own network, so that each user can (i) choose a suitable operator for each interface, and (ii) choose an interface for each active application. We make the following contributions in this paper. First, we propose concise network signalling that assists users to make informed choices even under mobility. Second, we develop user-choice algorithms that maximize a suitable notion of user satisfaction while using spectrum resources efficiently. Third, we perform extensive evaluation over actual base station deployment in a city coupled with real signal propagation maps. Our results with two operators show that, MOTA service model provides capacity gain in the range $2.5 \times -4\times$ over the current existing service model. Finally, we argue that our solution is practically implementable

by combining appropriate IEEE standards and IETF proposals.

Categories and Subject Descriptors: C.2.0 [General]: Data Communications; C.2.1 [Computer Communication Networks]: Network Architecture and Design-*Wireless communication*

General Terms: Design, Algorithms

1. INTRODUCTION

In the mobile data world, there are two interesting trends that are gaining attention of industry and media pundits. First, mobile data is growing at a rapid rate with several network equipment vendors (Cisco, Alcatel-Lucent, Ericsson, etc.) predicting that mobile data traffic will grow anywhere between $25 - 50\times$ by 2015 [11]. The FCC has taken note of this data tsunami, and in its National Broadband Plan [5], the FCC has called for *(i)* expedited release of 500 MHz of additional spectrum for mobile broadband and *(ii)* technical and business innovations that increase efficiency of spectrum utilization. The second interesting trend is that users are demanding greater degree of choice in choosing their operators at fine timescales. For example, across Asia, dual-SIM phones have become extremely popular [4]. This allows users to switch between operators based on signal quality and price plans. Further, Apple is rumored to be innovating (in collaboration with Gemalto, a SIM card maker) on a novel concept called Universal SIM [3] that allows iPhone users to toggle between different operators at will. This latter trend points towards an impending disruption in wireless service models. In this paper, we argue that, a carefully designed system that provides users the flexibility to choose operators at fine timescales is also the need of the hour from the spectrum shortage perspective.

Another interesting aspect of today's network deployment is that different operators deploy their networks to optimize different performance metrics. This means that depending on location and application requirements, the user choice of best operator could be different.

This is also validated by measurement studies at Cambridge [8], which show that operators indeed deploy their networks differently resulting in different quality of coverage from each of their technologies (2G, 3G) at different locations in Cambridge. Another example is a recent test of the iPhone4 on AT&T's and Verizon's 3G networks, which showed that the Verizon network was better for coverage and voice quality, while the AT&T network was superior for data downloads [12].

Thus it is clear that if users have choice, they can po-

---

[*]This work was carried out when Kanthi Nagaraj was with Alcatel-Lucent Bell Labs India.

tentially reap the benefits of this diversity in operator deployments. The immediate question is: does it suffice for users to make intelligent choices without any additional assistance from the operator? Since the only network information available to a user today is the channel quality, a simple thought experiment illustrates that this information is not sufficient in the following scenarios: *(i)* Since operator deployments are non-uniform, in any crowded area (e.g., few 100 sqm), all users in that area will associate with the operator that provides best signal quality. This naive association policy can degrade user experience and exacerbate perceived spectrum scarcity. *(ii)* When a user is highly mobile (moving at 100 Kmph), current signal quality may not be reflective of future user experience. Performance degradation in both scenarios could be addressed by users performing frequent operator selection. However, this would require frequent inter-operator hand-offs resulting in significant switching overheads. Further it could result in a ping-ponging effect. Therefore allowing users to make decisions without any network assistance is not a good design strategy.

The MOTA Service Problem: Motivated by the above discussion, we propose the *Mobile Operator and Technology Agnostic Access* (MOTA) service model. In our model, users are equipped with mobile devices with multiple radio interfaces (e.g., WiFi, 3G, LTE), all of which can be simultaneously active, each over a different operator (e.g., AT&T, Verizon). The MOTA service model empowers users to: *(i)* choose a suitable operator for each technology interface and *(ii)* associate each application with a suitable technology interface. Within this service model, our goal is to design a system that maximizes user experience, efficiently utilizes spectrum resources and minimizes control overheads.

Design Challenges and Requirements: There are two critical challenges in achieving the above goals.

- *Distributed Decisions:* In MOTA service model, users must make distributed decisions for the following reasons. First, since operators are self-interested, they cannot be expected to exchange network state information (load, user locations etc.) with each other to facilitate appropriate user allocation to different operators. Second, we cannot expect a third party entity to collect all this information and make centralized decisions, since the computational complexity of centralized decision making can be easily overwhelming.

- *User Mobility:* Since operators do not deploy their networks uniformly across locations, the right choice of operator for a user not only depends on her current location, but also on her mobility path. Thus the choice of operator for a mobile user should also account for future anticipated experience from each operator.

To overcome these challenges, we make the following design choice: *operators broadcast concise control signals (i.e., it does not scale with number of users) that capture information about their networks. This is used by client devices to make informed operator choices.* The broadcast signals include information related to user experience (e.g., application performance, price etc.). Given this design strategy, we address the following:

1. *Network Signalling:* What information should be maintained and broadcast by each operator? Can informa-

tion about future user experience under mobility be conveyed in a concise manner?

2. *User Algorithms:* How can a user make use of the network signals to choose suitable operators for her network interfaces and suitable interfaces for her applications ?

3. *Practical Feasibility:* Can all this be implemented with minimal changes to current network architectures and standards?

Note that addressing the above questions calls for developing new techniques and algorithms. Though user choice algorithms exist in WiFi and cellular contexts, none of these apply to the MOTA model (see Section 8).

1.1 Our Contributions

1. *MOTA Service Model:* We propose a framework for MOTA that enables users to switch seamlessly between operators at will. See Section 2, 3.

2. *Network Signaling:* Towards engineering the MOTA service model, we propose concise network signalling that each operator has to broadcast. Our signalling accounts for user mobility. See Section 4 and 5.

3. *User Algorithms:* We design user algorithms that exploit network broadcast signals to make suitable choice of operator for each technology and suitable choice of technology for each application. When users are static our algorithms have bounded gap from the optimal. See Section 4 and 5.

4. *Extensive Evaluation:* We perform extensive simulations using real operator deployments and real RF maps to evaluate the performance of our algorithm. Our results show that depending on the mix of mobile and static users, the system throughput gains with MOTA over existing service models can range from over 2.5 − 4×. See Section 6.

5. *Practical Feasibility:* We argue that the MOTA service model is implementable over existing cellular architectures by exploiting the IEEE 802.21 media independent handover standard and MPA and FastMIPV6 IETF proposals.

2. SYSTEM MODEL

2.1 Terminologies

We first provide some terminologies that we will use throughout the paper and then describe the basic setting of our system.

Technology: By technology, we mean the radio access technology, e.g., 3G/LTE.

Radio Interface: By radio interface of a client, we mean the RF hardware along with the PHY and MAC layers associated with a particular technology. A client could be equipped with multiple radio interfaces, each corresponding to a different technology.

Base Station: By base station, we mean the network element that provides wireless access over a particular technology. A physical base station could host multiple wireless
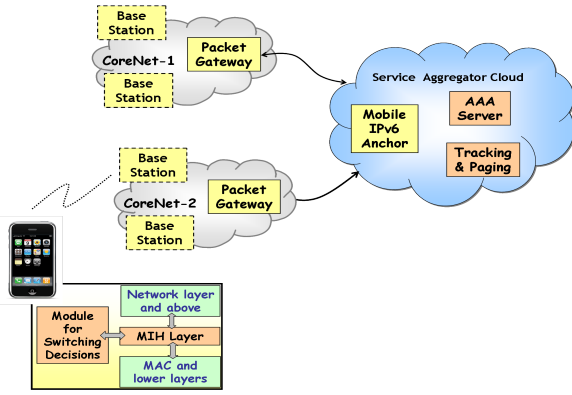
Figure 1: Functional separation between the operator's core network, service aggregator and client. Service Aggregator interfaces between operators and end users. The client's MIH (Media Independent Handover) layer is responsible for, *(i)* interpreting operator signalling information and passing it to the Module for switching decisions, and *(ii)* initiating signalling for inter operator switching.

technologies. However in this paper, a base station is the logical network element that offers access over only one specific technology.

Service Provider/Operator: In this paper, we say service provider/operator to refer to an entity that owns elements of wireless access network, including base stations, spectrum, backhaul infrastructure.

2.2 System Model

Each client device has multiple technology interfaces and it runs several concurrent applications. At each point in time, an interface can associate with only one operator offering that technology. Each client application can associate with only one radio interface[1].

The functional system model for the MOTA service is shown in Figure 1. To enable our service model, mobile users cannot have permanent relationships with operators. Thus there is a new intermediary entity called the *service aggregator*[2] responsible for maintaining permanent customer relationships. Service aggregator handles all control plane operations that cannot be handled by one operator alone. These operations include *(i)* billing and authentication, *(ii)* tracking and paging functionality that is used to locate users for push based services (this is handled by the Mobility Management Entity in LTE [36]) and *(iii)* seamless switching of user applications at Layer 3. All data plane operations and control plane operations that can be handled independently by a single operator remains under the purview of the operator. For e.g., these control plane operations include inter base station handover of the same operator, assignment of QoS classes, MAC scheduling etc.

REMARK 1 (SEAMLESS SWITCHING). *An important issue is how seamless switching is achieved (at Layer 3) when users*

---

[1]Our setting is extensible to scenarios where multi-homing of applications is allowed.
[2]This logical entity could be an operator, but logically separate from the entity that provides the access infrastructure

move from one operator to the other. We will discuss the feasibility of this in Section 7.

REMARK 2 (SWITCHING TIME SCALES). *Network conditions can be highly dynamic due to mobility, application churn, channel variations etc. Thus, an important issue is the frequency of switching decisions by clients. In this work we assume that switching decisions are made every T time units or when performance degrades significantly. T is chosen to trade-off spectrum utilization and signalling overhead. In our evaluation (Section 6) we will investigate this trade-off.*

3. THE MOTA FRAMEWORK

Recall that our design strategy is to allow users to make operator association decisions based on concise signalling information from the operators, so as to simultaneously *(i)* achieve high spectrum utilization, *(ii)* achieve QoS through service differentiation and *(iii)* maximize a suitable metric of user satisfaction. This will require us to answer the following questions:

- **Q1:** What information should each operator maintain?

- **Q2:** Based on the above, what aggregate information should be broadcast by each base station?

- **Q3:** What information should each user maintain?

- **Q4:** Based on the above two, how should a client make the following decision. *For each technology, which operator should it associate with, and to which technology interface should each application associate with?*

To answer these questions, we first need a formal framework that captures spectrum utilization, service differentiation and user satisfaction. Towards achieving high spectrum utilization, we use proportional-fairness which is a standard metric used in current cellular systems (EV-DO, 3G, LTE) to achieve a good trade-off between utilization and fairness [39]. Due to the vagaries of the wireless channel, providing hard guarantees will require significant over-provisioning of scarce radio resources. Therefore, for service differentiation, we use DiffServ like soft-QoS using weighted-GPS [33]. In soft-QoS model, there is a weight associated with each application class that reflects its relative priority. We capture user satisfaction via a utility function, which depends both on her throughput and the access cost.

Utility Framework: Standard proportional-fairness is equivalent to maximizing $\sum_a \ln R_a$, where R_a is the average throughput that application-a receives and the summation is over all applications. Towards providing soft-QoS, each application has a weight w_a associated with it. Thus maximizing $\sum_a w_a \ln R_a$ simultaneously achieves proportional fairness and service differentiation [34][3]. Note that base stations need to know the weights in order to achieve service differentiation through scheduling. In cellular networks like LTE, base station schedulers can determine this with the aid of an entity called the service gateway [36].

In our framework, we also allow each operator to charge a price for each technology. Since the weight of an application

---

[3]As an illustration, suppose application 1 and 2 are associated with same base station and experience identical channel conditions, then our objective function ensures the guarantee given by weighted-GPS, i.e., $R_1/w_1 \approx R_2/w_2$.

reflects its relative priority, we assume that the price (per unit time) incurred by a user is proportional to the weight of the class the application belongs to. More precisely, base station-j charges an amount $p_a = P_j w_a$ per unit time for an application with weight w_a. Note further that P_j is operator specific and does not change from one base station to the other of the same operator and technology.

We can now define a suitable metric to capture user satisfaction. Each user has two goals: on one hand, maximize $\sum_a w_a \ln R_a$ across her applications, and on the other hand, minimize the price she has to pay. There can be a tension between these two conflicting goals because the base station offering the highest data rate at any point could also be the most expensive one. To this end, we define each application's "utility function" as

$$\mathcal{U}_a(R_a, p_a) = w_a \ln(R_a) - \lambda_a p_a , \qquad (1)$$

where, R_a is the average throughput the application attains, p_a is the price per unit time the application pays depending on its association, w_a is the weight corresponding to the application class the application belongs to, and λ_a is a user specific constant (same for all applications of the user) that reflects the relative priority the user places on minimizing price as compared to maximizing the throughput. See Remark 3 on choice of λ. Such a utility function was first considered in [28] and subsequently used in many other contexts. Note that in our context, the slight difference is that our price is per unit weight and not unit rate. This is reasonable because there is no notion of fixed capacity in wireless systems and the weight reflects the average fraction of resources that a user is allocated from a base station. This utility simultaneously captures three aspects: proportional-fair rates, soft service differentiation, and price paid by a user. Indexing the applications by a, our goal is to maximize the following objective:

$$\max_{\{a:(R_a, p_a)\}} \sum_a \mathcal{U}_a(R_a, p_a) \qquad (2)$$

The achieved throughput R_a depends on the technology application associates itself to, the offered PHY-layer data rate, the load on the network etc.

REMARK 3 (CHOICE OF λ:). *The parameter λ reflects how much a user is willing to pay for additional throughput. To understand this, suppose the throughout that a user gets increases by a factor $\alpha > 1$, then the user's net utility increases if the price P_j of the associated base station increases by no more than $\ln(\alpha)/\lambda$. Therefore, if λ is high, users are not willing to pay too much for additional throughput and vice versa.*

REMARK 4 (ENERGY CONSIDERATIONS:). *Since we allow users to use multiple interfaces simultaneously, energy drain could be a concern. Our utility framework can account for this also, by setting an energy price for each interface based on current battery levels.*

Note that the price sensitivity and the energy sensitivity can be captured through parameters that are local to users. In our model and algorithms, the network need not be aware of these parameters.

Table 1: List of symbols used

Symbol	Description
$\mathcal{L}$, l ($l \in \mathcal{L}$)	Set of technologies and index for a typical technology, respectively
$\mathcal{J}$, j ($j \in \mathcal{J}$)	Set of base-stations and index for a typical base-station, respectively
$\mathcal{J}_l$	Set of base-stations offering technology-l
P_j	Price (per unit time per unit weight) charged by base-station-j
$\mathcal{J}_l^o$	Set of base-stations of operator o offering technology-l
$\mathcal{A}$, a ($a \in \mathcal{A}$)	Set of all applications and index for a typical application, respectively
$\mathcal{A}_u$	Set of applications of user-u
r_{uj}	PHY-layer data rate between user-u and base-station j
R_a	Throughput achieved by application a
w_a	Weight of application a (for service differentiation)
B_a	Duration of application a

3.1 Problem Statement

Our basic model is that, each client has multiple radios, each corresponding to a technology. There could be multiple applications running on one client. In the following, we will describe the system parameters, the constraints to keep in mind for solving the association problem, followed by the precise problem statement.

System parameters: The important system parameters are shown in Table 1. A typical user will be indexed by u. We will also denote by $\mathcal{A}$ the set of all applications, and a typical application will be indexed by a. Also, we denote by $\mathcal{A}_u$ the set of applications belonging to user u. We will denote by w_a the weight of the application class that application a belongs to. In practice, w_a takes values from small finite set. $\mathcal{J}$ is the set of all base stations. Each element $j \in \mathcal{J}$ belongs to some technology and some operator. Finally, $\mathcal{L}$ denotes the set of technologies available and $\mathcal{J}_l \subseteq \mathcal{J}$ denotes the set of base-stations offering technology $l \in \mathcal{L}$.

System constraints: Let x_{aj} be a variable that is set to 1 if and only if application a is associated to base station j. We now describe the three important system constraints for performing the associations.

1. *Integral association:* We assume that an application is associated with a *single* technology or base-station at a time, i.e., we do not consider multi-homing. This constraint essentially states that x_{aj} takes values from the set $\{0, 1\}$.

2. *Single operator per technology:* Clearly, at any given time, a radio interface can connect to only one operator. Let y_{uj} be a $0-1$ variable that is set to one if and only if $x_{aj} = 1$ for some $a \in \mathcal{A}_u$. Then, for user u and technology l, this constraint can be stated mathematically as

$$\sum_{j \in \mathcal{J}_l} y_{uj} \leq 1, \ \ \forall u, l . \qquad (3)$$

The association problem for MOTA: Define R_a as the average throughput application a receives. With $\mathcal{U}(.,.)$ as defined in (1), the association problem can be stated as follows:

$$\text{Maximize}_{x_{aj}} \quad \sum_{a \in \mathcal{A}} E[\mathcal{U}_a(R_a, p_a)]$$

subject to,

$$\forall \, u, \, l \, : \qquad \sum_{j \in \mathcal{J}_l} y_{uj} \leq 1$$

$$\forall u \, : \qquad \sum_{a \in \mathcal{A}_u} \sum_{j \in \mathcal{J}} x_{aj} w_a P_j = p_a$$

$$\forall \, a \in \mathcal{A}_u, \, j \in \mathcal{J}, \, \forall \, u \, : \qquad y_{uj} \geq x_{aj}$$

$$x_{aj}, \, y_{uj} \in \{0, 1\} \qquad (4)$$

The objective function is an expected value since network conditions are dynamic due to user mobility, application churn etc. The first constraint is explained before, the second constraint captures the price p_a the application-a has to pay, the third constraint ensures that y_{uj} is set to 1 if any application belonging to user u associates with j.

Accounting for Minimum Throughput Constraints: Though we have ignored minimum throughput requirements of some applications for ease of exposition, this constraint can be imposed by users excluding base stations not satisfying their requirements.

4. SIGNALLING AND ALGORITHM FOR STATIC CLIENTS

To address the four questions raised in Section 3, we first consider the case where all users are static. This is an important scenario to consider for two reasons. First, a large fraction of wireless data traffic is from nomadic users. This is validated by the fact that over 42% of iPhone data traffic was over WiFi [1], which is used primarily by stationary users. Second, eliminating the complexity of mobility allows us to build a solution in a structured fashion. In this section, we propose a base station signalling and a user algorithm with provable guarantees for static users and in Section 5 we develop a solution for the mobile case. See Section 5.1 for how these two solutions are combined.

Base Station Signalling: First, we make the following observation based on our assumption that all wireless technologies perform scheduling decisions that ensure proportional fairness and soft QOS guarantees[4].

FACT 1. *If the total weight of the applications associated with a base-station is W_j, then the throughput achieved by a client with weight w_a and physical layer data rate r_a under weighted-proportional fairness is $w_a r_a / W_j$.*

Fact 1 shows that, if a base station broadcasts the current total weights W_j of all applications associated with it, then all a user needs to do is estimate her PHY data rate (using channel measurements) to estimate throughput.

NETWORK SIGNALLING FOR STATIC USERS. *Each base station j transmits the load W_j and the price P_j.*

User Algorithm: For any user-u, the association problem can be stated as follows: associate each application to a suitable base-station, subject to *integral association* and *single operator per technology constraint* (see Section 3), such that

---

[4]WiFi does not provide proportional fairness, but there are a number of solutions to address this in WiFi

we maximize $\sum_{a \in \mathcal{A}_u} \mathcal{U}(R_a, p_a)$. Note that choosing one base station for an interface is equivalent to choosing the operator for that interface, since each base station is associated with one operator. A naive brute force approach for solving the above could be computationally heavy for the following reason. Assume that user u has K radio interfaces, J base station choices per interface and A_u active applications. Then the complexity of a brute force approach is $O(K^{A_u+J})$. Therefore even if the user runs only a few applications (say 5), the complexity of figuring out the optimal association can blow up. In fact, this problem is NP-hard[5]. Therefore we devise an algorithm with a constant factor approximation guarantee.

Recall that, in our scheme, each base-station broadcasts its current weight, based on which the user performs the association. Using Fact 1, the throughput of application-a can be expressed as

$$R_a = \sum_j \frac{x_{aj} w_a r_{uj}}{W_j + \sum_{a \in \mathcal{A}_u} x_{aj} w_a} \, ,$$

because, the association of u's applications to base station-j increases its current weight to $W_j + \sum_{a \in \mathcal{A}_u} x_{aj} w_a$. Clearly, only one of the $0-1$ association variables x_{aj}'s can take value one, i.e., $\sum_j x_{aj} = 1$. Using the above expression for R_a, we now note that,

$$\sum_{a \in \mathcal{A}_u} \mathcal{U}(R_a, p_a) = \sum_{a \in \mathcal{A}_u} w_a \left[\ln\left(\sum_j \frac{x_{aj} w_a r_{uj}}{W_j + \sum_{a \in \mathcal{A}_u} x_{aj} w_a}\right) - \lambda_u \sum_j x_{aj} P_j \right]$$

$$= \sum_{a \in \mathcal{A}_u} w_a \ln\left(\sum_j \frac{x_{aj} r_{uj} e^{-\lambda_u P_j}}{W_j + \sum_{a \in \mathcal{A}_u} x_{aj} w_a}\right) + \sum_{a \in \mathcal{A}_u} w_a \ln w_a$$

Since the second quantity in the above is devoid of optimization variables x_{aj}'s, maximizing $\sum_{a \in \mathcal{A}_u} \mathcal{U}(R_a, p_a)$ is equivalent to maximizing

$$V(\{x_{aj} : a \in \mathcal{A}_u, j \in \mathcal{J}\}) = \sum_{a \in \mathcal{A}_u} w_a \ln\left(\sum_j \frac{x_{aj} s_j}{W_j + \sum_{a \in \mathcal{A}_u} x_{aj} w_a}\right) \, ,$$
$$(5)$$

$$\text{where,} \qquad s_j \stackrel{\triangle}{=} r_{uj} e^{-\lambda_u P_j} \, .$$

Although s_j depends on both u and j, we have hidden the dependence of s_j on u since we are interested in u's decisions. In the following, we will describe an algorithm to find x_{aj}'s that maximize the expression given by (5).

Algorithm MOTA-Static: Our algorithm builds on the observation that, if we knew the total weight of all applications associated with a particular radio interface, then, it is straight-forward to determine the best base station for that interface. We then use this observation and run a greedy algorithm to select an operator for each radio interface, and select an interface for each application.

1. *Sub-routine for base station selection for a technology:* Suppose w is the total weight of all applications of user u that gets associated to base station j, i.e., $w = \sum_a x_{aj} w_a$. Then, it is easy to see from (5), that the total contribution of base station j to $V(\{x_{aj} : a \in \mathcal{A}_u, j \in \mathcal{J}\})$ is given by

$$w \ln\left(\frac{s_j}{W_j + w}\right) \, .$$

---

[5]This follows by reducing the problem in [30] to our problem.

Algorithm 1 MOTA-STATIC: Assignment of applications to radio interfaces and operators to interfaces.

1: *Initialization:* Order the weights of the applications of user u as $w_a, i = 1, 2, \ldots, n$, where $w_1 \leq w_2 \leq w_3 \leq \ldots$. Let Δ_l be the total weight of all application associated with interface l. Set $\Delta_l = 0, \forall l$.
2: **for** $r = 1, \ldots, n$ **do**
3:   Assign application r with weight w_r as follows:

$$l_r = \arg \max_l \left[(\Delta_l + w_r) G_l(\Delta_l + w_r) - \Delta_l G_l(\Delta_l) , \right]$$

   where, $G_l(\cdot)$ is defined by 7.
4:   $\Delta_{l_r} = \Delta_{l_r} + w_r$
5: **end for**
6: Return $j_l^*(\Delta_l)$ for all l

However, we have a constraint that states that only one base station can be selected from a technology. From the preceding we can see that, this selection depends on the total weight w, of all applications of user u that we associate with the technology. In particular, for technology l, the best base station would be given by

$$j^*{}_l(w) = \arg \max_{j \in \mathcal{J}_l} \ln(\frac{s_j}{W_j + w}) , \qquad (6)$$

where we have explicitly shown that dependence of $j_l^*(w)$ on w, the total weight of all applications of user u that is associated with technology l. To this end, for each technology l, we define the function $G_l(w)$ as

$$G_l(w) \triangleq \max_{j \in \mathcal{J}_l} \ln(\frac{s_j}{W_j + w}) . \qquad (7)$$

This sub-routine thus computes the function $G_l(w)$ for a given w and also the corresponding best base station $j^*{}_l(w)$ given by (6).

2. *Greedy application association:* The MOTA-Static algorithm is described in Algorithm 1.

In the initialization step 1, we first order the applications of user u in increasing order of weights. Further, we define a quantity Δ_l, the total weight of all applications assigned to interface l and initialize this to zero. Next we iterate through each application in increasing order of weight. Step 3 is the greedy step, where we associate application-r to that interface l_r which results in the maximum increase in utility over the current asociation of the first $r-1$ applications. Step 4 merely updates the total weight of the appropriate interface. Once all the applications are assigned interfaces, we use equation (6) to determine the best base station for each interface.

Performance guarantee of Algorithm MOTA-Static: It can be shown that the greedy algorithm described above has good performance guarantees, especially when the base stations have a large amount of traffic. Specifically, consider an instance where base station-j broadcasts its total weight of all associated applications as W_j. Suppose, user u wishes to associate n applications with weights $w_1, w_2, \ldots, w_n$ to different base stations using Algorithm MOTA-Static described earlier, and also suppose $w_{sum} = \sum_{a=1}^{n} w_a$. Suppose α is the maximum load all applications of user-u applications can together contribute to any base station, i.e.,

$$\alpha = \max_j \frac{w_{sum}}{W_j + w_{sum}}$$

The following holds for Algorithm MOTA-Static.

THEOREM 1. *Let $\mathbf{r}_u = (r_{u1}, r_{u2}, \ldots)$ denote the vector for the data rates of user u to the different base stations. Let $OPT(\mathbf{r}_u)$ and $ALGO(\mathbf{r}_u)$ be the total utility of an optimal algorithm and our MOTA-Static algorithm, respectively, as a function of $\mathbf{r}_u$. Then,*

$$ALGO(\mathbf{r}_u e^{2\alpha}) \geq OPT(\mathbf{r}_u) ,$$

where $\alpha \in (0, 1]$ is the maximum load all applications of u can contribute to any base station.

The proof can be found in the longer version available at [7]. It essentially relies on the fact that the function $G_l(w)$ is concave in w.

REMARK 5. *Two interesting cases to consider are when α is very small and large. When α is very small, the performance of our algorithm gets provably very close to the optimal. This is seen by setting α to zero in Theorem 1. This is the typical operating regime when all available base stations are loaded heavily. The other interesting case is when $\alpha = 1$ which arises when the base station is a Femto solely owned by the user. In this case, our algorithm is an e^2 approximation in the worst case.*

4.1 Price of Anarchy and System Efficiency

We have proposed a strategy where each user *selfishly* associates its applications without considering the effect on other applications of other users. In general, such a design can lead to a scenario where the eventual solution is arbitrarily worse off than the optimal, i.e., the *price of anarchy*[6] is unbounded. Few examples are provided in [35] in the case of selfish routing games. The question is: is there such a possibility in our association strategy? The following theorem shows that the answer is no. Essentially, the proportional fair allocation by each base station based on weights ensures that such a possibility cannot arise. The following theorem captures this result.

THEOREM 2. *Consider a setting with all static users. Let $GLOBAL(\mathbf{r})$ and $SELFISH(\mathbf{r})$ respectively be the global optimal utility and the total utility of our selfish strategy as a function of the data rates of all the users to all the base stations (denoted by $\mathbf{r}$). Then, there is a universal constant $1 \leq K < \infty$ (independent of all network parameters) such that*

$$SELFISH(K.\mathbf{r}) \geq GLOBAL(\mathbf{r}) .$$

The proof is in the longer version available at [7]. The above result simply shows that the performance of our strategy has bounded gap from optimality. While this is an important aspect of our design, our evaluation demonstrates the performance with a mix of static and mobile users.

5. SIGNALLING AND ALGORITHM FOR MOBILE CLIENTS

In the previous section, we designed base-station signalling and user application association algorithms when a user is static. However, if users move around from one cell to another, the situation is more complex. If mobile users could

---
[6]The term *price of anarchy* is used in game theory to describe the gap from optimal when users take local selfish decisions.

make association decision for their applications at every instant (or at very short intervals), then we could apply our algorithm for static users at those instants. However, as discussed in Remark 2, association decisions can only be taken periodically to keep the signalling overhead of association switching low. Thus, it is imperative that, association decisions of applications under mobility take into account the "expected" application performance over one association period (over which users could move across multiple cells).

Our goal is to design base station signalling and user-end application association algorithms so as to maximize the overall expected utility of all applications. In the mobile case, once a switching decision is made, the utility received by an application is not only a random variable, but it also varies as a mobile user is handed-off from the current cell (the cell where an association decision is made) to another cell of the operator. The identity and the state (load) of the next cells on the user mobility path are random variables from the point of view of the user module that makes the association decisions. Thus, we maximize $\sum_a E[\mathcal{U}(R_a, p_a)]$ (summation is over all applications of all users), where $\mathcal{U}(R_a, p_a)$ denotes the time-average utility for application-a.

We will now describe our proposed base-station signalling and user algorithm.

Base Station Signaling: We wish to derive the information that base-stations should broadcast that assist users to make association decisions based on anticipated user experience. Towards this end, we first answer the following question:

Q-EU: *What is the expected time-average utility of application a of a mobile user u that associates application-a (with weight w_a) to base-station-j of operator o?*

The system is inherently much more complex than the one with static users due to the randomness introduced by user mobility pattern, user behavior and application churn[7]. Thus, to answer Q-EU, we make use of the following practical property:

P1: *Every base station carries much more load than the total weight that applications belonging to any one user can contribute. In other words, for every j, $W_j \gg \sum_{a \in \mathcal{A}_u} w_a$ where $\sum_{a \in \mathcal{A}_u} w_a$ is the total weight of all applications of user u. Note that, this property is not reasonable in the static scenario as users have options of associating with Femto base stations and Wi-Fi access points that are not necessarily heavily loaded.*

To answer Q-EU, consider a technology interface l. Note that, once user-u chooses a particular base station for application a, then as the user moves, all hand-offs are to base stations belonging to the same operator. Let $c(t)$ be the base station of operator o that user u associates with at time t. Let $W(t) = W_{c(t)}$, the total weight of the base station that provides coverage to user u at time t. Clearly, $W(0) = W_j$ since a initially associates with base station j. Also, let $s_{u,k} = r_{u,k}e^{-\lambda_u P_j}$, and $s_u(t) = s_{u,c(t)}$, $r_u(t) = r_{u,c(t)}$. Note that we assume $P_{c(t)} = P_j$ depends only on the operator of base-station, i.e., it does not change across base stations of the same operator.

Now note that,

$$w_a \ln \tfrac{s_u(t)w_a}{W(t)+w_a} \approx w_a \ln \tfrac{s_u(t)}{W(t)} + w_a \ln w_a \quad \text{(using \textbf{P1})}$$
$$= w_a g_{j,a}(t) + w_a \ln w_a , \tag{8}$$

where, $g_{j,a}(t) = \ln \tfrac{s_u(t)}{W(t)}$. Since $w_a \ln w_a$ does not depend on the associated base station, we will only deal with the random quantity $g_{j,a}(t)$. Denoting the time average of $g_{j,a}(t)$ over the switching duration by $\overline{g_{j,a}}$, we derive $E[\overline{g_{j,a}}]$ as follows:

$$E[\overline{g_{j,a}}] = E\left[\tfrac{1}{T}\int_0^T g_{j,a}(v)\ dv\right]$$
$$= E\left[\tfrac{1}{T}\int_0^T (\ln(r_u(v)) - \ln(W(v)))\ dv\right] - \lambda_u P_j$$

Since the user moves from one cell to another after starting from cell j, intuitively speaking, the integral in the above time-average can be computed by adding up contributions from all the cells of operator o visited by use u. We will make this intuition precise in the following. Define $T_{j,k}$ as the following random variable:

$$T_{j,k}(a) \triangleq \begin{array}{c} \text{Time spent in cell-}k\text{ by} \\ \text{user-}u\text{'s application } a \text{ intiated in cell-}j \end{array}$$

$$R_{j,k}(a) \triangleq \begin{array}{c} \text{Aggregate } \ln(r_u(t)) \text{ over the time} \\ \text{spent in cell-}k \text{ by user-}u\text{'s application } a \\ \text{that is intiated in cell-}j \end{array}$$

Then, denoting application duration by B_a, we write

$$E[\overline{g_{j,a}}]$$
$$= \tfrac{E[\text{Aggregate } (\ln(r_u(t)) - \ln(W(t))) \text{ over } a\text{'s lifetime}]}{EB_a} - \lambda_u P_j$$
$$= \tfrac{\sum_{k \in \mathcal{J}_l^o}\left(E[R_{j,k}(a)] - (\ln W_k)E[T_{j,k}(a)]\right)}{EB_a} - \lambda_u P_j \tag{9}$$

Note that, in the second line, we have rewritten expectation of a fraction as the fraction of expectations. The rationality of this comes from Markovian reward theory [38]. In Markovian renewal reward theory, a Markovian process (i.e., a future state of the process only depends on the past through the current state) collects rewards that depends on the state of an underlying Markovian process. The fundamental theorem of reward theory says that, under suitable assumptions including stationarity, expected time-average reward is equal to, expected reward over a reward cycle divided by the expected duration of the reward cycle. The process $g_{j,a}(t)$ can be viewed as a renewal process by "stitching" together in time all instances of application-a starting when the user of a is in cell-j. However, assumption of stationarity of the renewal process (for applying reward theorem) does not hold in our case[8]; we have nevertheless used intuition from reward theorem to derive expression given by (9).

The expression given by (9) immediately reflects the information that has to be maintained by base-station j for mobility support:

- $\sum_k E[R_{j,k}(a)]$: For each cell j, each *operator* maintains estimates of $\sum_k E[R_{j,k}(a)]$ for each *application class* using historical data. In practice, maintaining this value for four [9] (data, video, voice, gaming) should suffice. Base station k can maintain historical estimate

---

of $E[R_{j,k}(a)]$ and the operator aggregates this data to compute the desired quantity for each j. Estimation of $E[R_{j,k}(a)]$ at each base-station can be done as follows. For each application a of user u that was started in j and that enters k while alive, base station k computes the aggregate of $\ln(r_{u,k}(a))$ over the time spent in k based on the channel state information at discrete time intervals. Each sample collected from an application can be used to compute the estimate of the desired expectation for the application class a belongs to.

- $\sum_k \ln(W_k)E[T_{j,k}(a)]$: The operator also maintains this quantity for each base station j and each application class. Note that, W_k is the current weight, and $E[T_{j,k}(a)]$ is based on historical data as described above. Again, each cell k can periodically update the operator about current W_k and estimated value of $E[T_{j,k}(a)]$.

The network signalling is described below.

NETWORK SIGNALLING FOR MOBILE USERS. *Each base station j should broadcast the quantity $\sum_k E[R_{j,k}(a)] - (\ln W_k)E[T_{j,k}(a)]$ for each application class separately, and the price P_j. Here W_k's are based on current weight and the expectations are estimated using historical data as described above.*

For each application a, the user, upon obtaining the broadcast information, uses a local estimate of EB_a to compute $E[\overline{g_{j,a}}]$ in (9), which plugged into (8) gives the desired expected utility.

User Algorithm: We now turn to the problem of how a user determines the following: *(i)* which set of applications to associate with each technology and *(ii)* which base-station to select for each technology. These are obviously coupled problems. Note that, each base station corresponds to an operator and so, by choosing one base station for a technology interface we are essentially choosing one operator for the technology interface. Consider a user u who wishes to perform the optimal association. Let $U_j(a)$ be the expected utility that application a achieves if user u associates application a to base-station j (of some operator). In the previous subsection, we described how $U_j(a)$ can be computed based on base-station signalling. Suppose x_{aj} is the $0-1$ variable that is set to one if a is associated to base-station j. Then, our goal is to maximize $\sum_{a \in A_u} \sum_j x_{aj}U_j(a)$ subject to the constraint that only one operator can be chosen from a technology. Note that, this problem is different from the static case for the following reason. In the static case, if multiple applications were associated to one base-station, they would get utility proportional to their weight. We exploited this structure to derive Algorithm MOTA-Static in the static case. Unfortunately, there is no such structure in this mobile case as the expected utility of an application has complex dependence on its duration and the cells it is handed-off to subsequently. Nevertheless, the problem in mobile case of maximizing $\sum_{a \in A_u} \sum_j x_{aj}U_j(a)$ is similar to the problem of maximum generalized assignment problem (GAP) [23] of assigning balls (applications) to bins (technologies). The main difference in our problem is that we have no constraints on the set of items (in this case applications) that can be assigned to an interface. Further there is an additional element of having to choose an operator for each interface. We suitably modify the local search algorithm in [23] as below.

Algorithm 2 MOTA-MOBILE: Assignment of applications of user u to radio interfaces and base-stations (operators) to interfaces.

1: *Initialization:* Assign every application to a random interface and pick a random base station for every interface. Let S_{lj} to be the set of applications of user u that get associated to operator j. Let $v(a)$ be the marginal utility of the application a. The $0-1$ variable $x_{aj} = 1$ iff application a is associated to base station j, and, the $0-1$ variable $z_{al} = 1$ iff application a is associated to some base station of technology l.
2: **for** $L \ln \frac{1}{\epsilon}$ iterations **do**
3:     **for** $l \in \mathcal{L}$ **do**
4:       **for** $j \in \mathcal{J}_l$ **do**
5:         Construct $S_{l,j}$ and compute Δ_{lj} as follows:

$$S_{l,j} = \{a : \text{either } z_{al} = 1 \text{ or } \delta_{lja} > 0\},$$
$$\Delta_{lj} = \sum_{a \in S_{l,j}} \delta_{aij}$$

6:       **end for**
7:     Compute

$$\Delta_l^* = \max_{j \in \mathcal{J}_l} \Delta_{lj}, \qquad j^*(l) = \arg\max_{j \in \mathcal{J}_l} \Delta_{lj}$$

8:     **end for**
9:     $l_{best} = \arg\max_{l \in \mathcal{L}} \Delta_l^*$.
10:     For all $a \in S_{l_{best},j(l_{best})}$, set

$$x_{aj^*(l_{best})} \leftarrow 1, \quad v(a) \leftarrow U_{j^*(l_{best})}(a) \ ,$$

    and update z_{al}'s appropriately.
11: **end for**
12: Return the association decisions.

- *Initialization (Step 1):* First, all interfaces chose an arbitrary operator and each application chooses the interface whose choice of operator provides best expected utility.

- *Local search iteration (Step 2-11):* For each application a, let $v(a)$ be the current utility in an iterative step. Recall that x_{aj} is the $0-1$ variable that takes value one if a is associated to base station j through a suitable interface. A subset of the x_{aj}'s are updated after each iteration. For convenience, also define the variable z_{al} as the $0-1$ variable that is one iff application a is associated to some base-station offering technology l (i.e., $z_{al} = 1$ if for some $j \in \mathcal{J}_l$, $x_{aj} = 1$). Given an association from the previous iteration, we define the marginal utility of the application a if it were assigned to interface l and operator j over its current allocation as follows for all $j \in \mathcal{J}_l$:

$$\delta_{lja} = \begin{cases} U_j(a) - v(a), & \text{if, } z_{al} = 0, \\ U_j(a), & \text{if, } z_{al} = 1 \end{cases} \quad (10)$$

For all l,, construct the sets $S_{l,j}$ as follows:

$$S_{l,j} = \{a : \text{either } z_{al} = 1 \text{ or } \delta_{lja} > 0\} \ .$$

Let $\Delta_{lj} = \sum_{a \in S_{l,j}} \delta_{lja}$. Here, Δ_{lj} denotes the increase in utility if some applications are re-associated to base station j of technology interface l from some other interface. Let $\Delta_l^* = \max_{j \in \mathcal{J}_l} \Delta_{lj}$ and $j^*(l) = \arg\max_{j \in \mathcal{J}_l} \Delta_{lj}$. Base station $j^*(l)$ is the best base station of technology l for maximizing marginal utility in this iteration. Each iteration of local search is now easy to describe: based on the association from the previous iteration, obtain the technology interface l_{best} with maximum Δ_l^*, and,

on that interface update x_{aj} for $j^*(l_{best})$ and also update $v(a)$.

REMARK 6. *The analysis of local search algorithm in [23] goes through with minor changes in our case and the following result can be shown. Suppose, we start with an arbitrary initial association. Then, Algorithm 2 achieves at least $1/2 - \epsilon$ of the improvement possible over the initial association, with respect to the expressions of expected utilities used in this section.*

5.1 Putting it All Together

Each base station transmits signalling information developed in the static case as well as the mobile case. Note that the only additional piece of information required over the mobile case, is the load in the base station. An application on the user's phone determines whether it is static or mobile (either via GPS/accelermometer readings or coherence time measurements) to determine which piece of information to use. Some access points such as femto base station may only be able to provide static signalling.

6. EVALUATION

We perform detailed simulations to evaluate our algorithms using topology and RF maps from a real deployment. Our goals are twofold *(i)* demonstrate the benefits of the MOTA over existing policies for realistic user behavior *(ii)* study the impact of different parameters including switching frequency, user mobility, traffic mix etc. on system performance.

6.1 Evaluation Framework

Network Topology: We obtained cell tower information (including antenna heights, locations etc.) of a leading operator in a major city in Asia along with terrain information in that city[9]. Actual RF maps were generated by plugging in the tower and terrain information and technology (e.g., HSDPA, LTE) into a commercially available RF tool that is used by operators for cellular planing [2]. This information is extremely hard to obtain and we managed to coax only one operator to share this. However we need information for at least two operators. Fortunately, in Asia tower infrastructure is provided by third party entities and thus it is a common practice for operators to share tower infrastructure. Therefore, we assume that two operators deploy the same technologies at each tower location. Intuitively, MOTA will achieve larger gains when there is more diversity in operator deployment. Thus, the above scenario, provides a pessimistic evaluation for MOTA.

We consider a 5Km $\times$ 5Km area in the city. We assume that there are two operators. We assume that each cell tower has a HSDPA and LTE base station from each operators. We use the RF planning tool [2] to generate the signal quality (and hence technology dependent PHY data rate) at different locations within the 5 Km $\times$ 5 Km area. Figure 2 provides an example of signal strength predictions from the tool.

Application Models We divide the applications into three classes: voice, video and data. We generate each of these applications on a client using the models given in the evaluation guidelines for next generation mobile networks [9]. The weights for providing service differentiation is set to 1, 2 and 4 for voice, data and video respectively. Prices for all technologies are set to be equal.

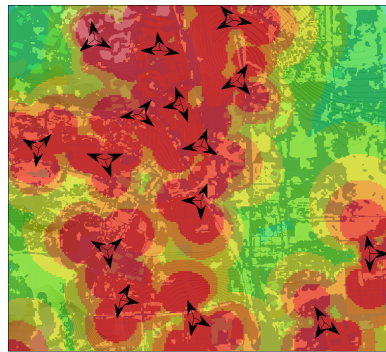[9]Due to an NDA we cannot reveal the operator and city name.

Figure 2: Signal strength map from RF tool

User Mobility: We simulate the mobility pattern of mobile users according to the Manhattan model [13] and the Random Waypoint model. Each user is equipped with LTE and 3G radio interfaces. Hand-offs are done based on signal-strength.
Performance Metric. We measure the average throughput per user and the utility function formulated in equation 1, to compare MOTA with the Naive scheme.

6.2 Results

We now present the results of our simulations. Our results are averaged over several runs with different random seeds. We run simulations for a duration of 20 simulation minutes.

Benefit of MOTA over Existing Service Models: We extend existing service models to multiple interfaces in a natural way as follows. Each user is associated only with a fixed operator, but multiple interfaces can be simultaneously active, with each interface associating with the base station of that technology with highest signal quality. We only show results when users are distributed in 60:40 ratio between the two operators; similar results were observed for a 50:50 mix. We compare this existing service model with the MOTA service model. We consider different mobility mixes, *(i)* all static users, *(ii)* 30% mobile users and *(iii)* all mobile users. Our results are shown in Figures 3(a) and 3(b).

Our main observations are as follows:

1. *Throughput Gains:* MOTA provides 2.5-4.0$\times$ gain in throughput. The benefits are the least (2.5$\times$), when all users are mobile. In other words, irrespective of mobility scenario, MOTA improves the spectral efficiency of the collective spectrum pool of mobile operators. The better gains in the static case can be explained as follows. MOTA-mobile relies on the predictive utility to pick the best base station while MOTA-static has complete information about the utilities of all visible base stations.

2. *Effect of Switching Duration:* Figure 3(b) shows that even when MOTA is run once in 180 s, performance degradation is negligible. Even if it is run once every 720 s, the performance degradation is about only 10% as compared to switching very frequently.

Benefit of MOTA over Optimized Single Operator Model: One might argue that the existing single operator
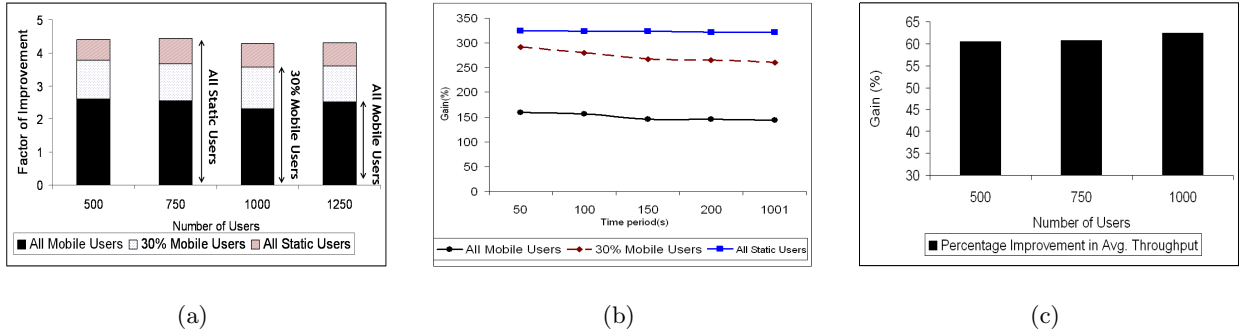
Figure 3: (a) MOTA gains over existing service models. (b) MOTA gains as a function of switching duration. (c) MOTA gains over optimized Single Operator.

service model, with users making more intelligent decisions on base stations based on load information etc. may be sufficient. To investigate this we conducted the following experiment. We skew the deployments of the two operators in terms of number of base stations that they provide by deploying 16 additional base stations for one of the operators. Users are distributed equally among the operators. Users choose the base stations that they connect to using the MOTA-Static algorithm, with the constraint that they can only choose from the operator they are associated with. We see that in this case, users gain more than 60% with the MOTA service model 3(c).

Benefit to Operators: While the MOTA service model clearly benefits end users and improves spectrum utilization, the question is what incentive do operators have to switch to this model? This a much broader problem which requires careful analysis. Nevertheless, we performed some preliminary experiments to see if there are benefits to operators. For the existing service model, we set $\lambda = 1$ for all users, $P_j = 1$ for all base stations and we evaluate the aggregate utility. In the MOTA model, we increase the price P_j's and obtain the price at which the MOTA model achieves the same utility as the existing service model. We observe that with all else remaining same, operators can set $P_j = 3$ with the MOTA service model, while providing the same total utility as the existing service model.

7. PRACTICAL FEASIBILITY

In this section, we will show how the MOTA service model is practically feasible by building on existing cellular architectures, IEEE standards and IETF proposals.

7.1 Background

3G and Beyond(3G+) Wireless Core Network Primer: In this paper we consider an all IP core network, since this is going to be true for all network architectures going forward (e.g., LTE [36])[10]. We discuss only the elements and their functions that are relevant to our subsequent discussions.

---

[10]We do not consider 2G, since countries are refarming 2G spectrum for 3G/4G use. For example Ofcom has already mandated this in UK.

1. *Packet Data Network Gateway (PGW):* The PGW is the interface between the operator's core network and external networks. The functionalities of PGW that are relevant for this discussion are *(i)* being a mobility anchor for inter technology (vertical) handovers and *(ii)* providing IP address to clients. The PGW is referred to as the GGSN in 3G and PGW in LTE.

2. *Mobility Management Entity (MME):* The MME (mobility management entity) is responsible for user authentication, tracking and paging of the user and selection of Service Gateway (SGW) and PDN Gateway (PGW).

IEEE 802.21 Primer: The IEEE 802.21 standard is a framework for media independent handover (vertical handoffs) between different wireless technologies (IEEE and 3GPP) [6]. The media independent handover function resides between Layer 2 and Layer 3 of clients and provides access technology independent interface and service primitives to Layer 3.

The framework allows for three types of vertical hand-offs, *(i)* terminal initiated and controlled, *(ii)* terminal initiated and network assisted (TINA) and *(iii)* network initiated and controlled. Clearly the MOTA service falls under the TINA model. While our work exploits the IEEE 802.21 framework, it goes well beyond its initial scope, by *(i)* allowing different applications to use different interfaces simultaneously and *(ii)* allowing users to choose among multiple operators even within the same interface. We now show how, MOTA service model can be enabled via IEEE 802.21 signaling and suitable modifications to the functionalities of certain 3G+ core network elements.

7.2 Operator/Technology Switching

During operator switching of a user interface, the key issue to address is seamless switching (i.e., make before break) and latencies due to authentication, connection establishment and network mobility (IP address change etc.). We will illustrate how this can be addressed by building on the IEEE 802.21 framework, combined with existing proposals in the IETF/IRTF. Refer to Figure 1.

As we can see from Figure 1, the following changes in functionality are required to the cellular architecture elements. AAA is moved entirely to the service aggregator. The tracking and paging functionality of the mobility management en-

tity is moved to the service aggregator. Further, the mobility anchor functionality played by the PGW is moved to the service aggregator. Base stations broadcast the signalling information described in Sections 4 and 5 using the signalling messages proposed in IEEE 802.21 for IEEE and 3GPP standards. We will now illustrate how switching between different operators is done, once a switching decision is made.

- *Step 0:* Suppose a switching decision has been made at a client. Let I be the set of interfaces that need to switch operators. Let A be the set of applications that are assigned a new operator, possibly by switching interface.

- *Step 1:* For each application in A and interface in I, the client sends a IEEE 802.21 handover initiation message to the service aggregator over any of the associated current networks.

- *Step 2:* For each interface in I, the service aggregator uses the IRTF's Media-Independent Pre-Authentication (MPA) framework [21], to acquire authentication, IP address and network resources (e.g., reserve resources or bearer paths between the PGW and Service Gateway) in the new operator's network.

- *Step 3:* Now for each application in A, the service aggregator use mechanisms such as Fast Handover in MIPv6 [10], to simultaneously establish a tunnel to the service gateway of the new network and forwards duplicate packets.

- *Step 4:* Once Step 4 is completed for all applications in A, the service aggregator sends a *IEEE 802.21 handover ready* message for each interface in I over the existing networks.

- *Step 5:* The client switches interfaces in I to the new network and switches the applications in A too.

Gathering Network State Information: In our service, network broadcasts some aggregate information that assists the user to make appropriate association choices. For this approach to work, each interface of a user needs to gather this information. The question is, if an interface of a technology is already associated with one operator, how does it overhear information from other operators of the same technology? We argue that this is possible with today's RF hardware. For example in today's cellular Frequency Division Duplex (FDD) systems, it is possible for a radio to transmit on the transmit frequency band and concurrently receive on the paired receive frequency band, so long as the paired bands are sufficiently separated. Thus when a user is transmitting to an associated operator over an interface, at the same time, she can receive broadcast information from another operator of the same technology over that interface (since these are different operators, there will be sufficient guard band between the two frequencies). From a hardware perspective such an approach can be used in Time Division Duplex Systems (TDD) also. For such an approach to work in TDD systems, the user occasionally switches to another operator's base station by becoming idle and not receiving anything from the associated base station. How frequently the user uses opportunistic overhearing mechanism will depend on switching overheads and packet losses. For example in LTE systems, a frame duration is 1ms, therefore if users opportunistically overhear control signals from other base stations every few 100 milliseconds, the overheads can be kept under 1%.

8. RELATED WORK

ABC and IEEE 802.21 The IEEE 802.21 standard emerged from the Always Best Connected (ABC) concept [25]. The ABC concept recognized the fact that mobile devices have access to multiple technologies (WAN, LAN, PAN) etc. The key goal of the ABC concept was to enable users to seamlessly switch to the best technology. For example, when a user walked into his home, his mobile would automatically switch to the Wi-Fi network since it provided better connectivity inside the home. A variety of research papers on ABC and IEEE 802.21 primarily focussed on seamless handover to the best available network to minimize latencies [24, 15, 20, 29, 32, 21, 22] Our work builds on top of this paradigm by *(i)* allowing simultaneous use of multiple interfaces, *(ii)* allowing interfaces to choose between operators and *(iii)* explicitly accounting for mobility in making informed choices.

One of the goals of the PERIMETER EU project (http://perimeter.tssg.org/) is to allow users to seamlessly switch between operators. In [18], the authors outline an architecture to enable users to seamlessly switch between operators and possible demo scenarios. However, the vision of this project is to enable users to seamlessly switch with no assistance from the operator. As we have shown, for optimal resource utilization, network assistance is essential. Further, we outline a precise architecture, base station signalling and user algorithms for optimal end user experience.

Related problems on user choice: The user choice problem in our work has two components, operator selection for different interfaces and interface selection for different applications. While we consider these coupled problems, special cases of these individual problems have been considered in a few contexts. For example, in Wi-Fi scenario, [27, 31] propose distributed algorithms and [30] proposes centralized algorithms for Wi-Fi clients to intelligently choose channels. Though the channel selection problem in Wi-Fi may look similar to operator selection in our case, there are two fundamental differences: we have to choose the operators for multiple active interfaces that are coupled through applications' choice of interfaces, and mobility of client is an important consideration for our setting. These aspects also call for developing new techniques. When the user choice is restricted to choosing one among multiple technologies, solutions are proposed [15, 37, 19]. However, none of these account for user mobility explicitly and these works do not extend to the case of multiple simultaneously active technology interfaces along with a choice of multiple operators within each technology. Finally, centralized algorithms for proportional fair user association within a single technology (single operator is implicit) have been developed in [14] for 3G and in [30] for Wi-Fi. Again none of these account for user mobility explicitly. Also, the techniques for centralized algorithm does not apply to our static-user case too as multiple operators in our setting calls for distributed algorithms.

Dynamic spectrum sharing: As opposed to MOTA approach of allowing user to make operator selection at fine time scale, the dynamic spectrum based approach allows operators to share spectrum through an entity called "spectrum broker [16, 17, 26]. We believe MOTA framework is a simpler approach for effectively sharing spectrum and considers user specific aspects like application requirements and mobility.

9. CLOSING REMARKS

We proposed and designed MOTA, a service model that

empowers users to choose the operator and technology. MOTA provides up to 4× capacity gains over existing service model, thus significantly improving the overall spectrum utilization. While the scope for further research in this space is huge, we believe that policy makers and technologist both have to come forward to make such a model a reality.

10. REFERENCES

[1] AdMob Mobile Metrics Report 2 (2008). See *http://www.admob.com/marketing/pdf/ mobile_metrics_nov_08.pdf* .

[2] Alcatel-lucent 9955 radio network planning tool. See *http://www.scribd.com/doc/40620661/9955-V6-9-RN-ED- 03.*

[3] Could Apple Be Your Next Phone Company?, available at *http://gigaom.com/2011/02/10/could-apple-be-your-next-phone-company/.*

[4] Dual-sim cellphone market takes off as tariffs slip, available at *http://economictimes.indiatimes.com/tech/hardware/Dual-sim-cellphone-market-takes-off-as-tariffs-slip-/articleshow/5526075.cms.*

[5] FCC's national broadband plan available at *http://www.broadband.gov.*

[6] IEEE 802.21 Working Group. See *www.ieee802.org/21/* .

[7] Mota: Engineering an operatior agnostic mobile service, tech. report., *http://www.bell-labs.com/user/vikramsr/papers/mota-longer.pdf.*

[8] Narseo Vallina-Rodríguez and Jon Crowcroft, Erdos: Enabling opportunistic resources sharing in mobile Operating Systems, ICDCN 2011, Keynote talk . See slide 55 *http://www.cl.cam.ac.uk/ jac22/talks/icdcn.pdf* .

[9] NGMN Radio Access Performance Evaluation Methodology available at *http://www.ngmn.org/uploads/media /NGMN_Radio_Access_Performance_Evaluation_Methodology.pdf.*

[10] RFC 4068-Fast Handovers for Mobile IPV6, available at *http://www.faqs.org/rfcs/rfc4068.html.*

[11] The Network Paradox Part I: Mobile Data Demand by the Numbers, available at *http://connectedplanetonline.com/4gparadox/Mobile-data-by-the-numbers-1005/index.html.*

[12] Verizon iPhone Gets Better Coverage, Slower Data, Wired.com Test Finds, available at *http://www.wired.com/gadgetlab/2011/02/ verizon-att-iphone-tests/.*

[13] BAI, F. SADAGOPAN, N. H. A. Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks. In *IEEE INFOCOM.*

[14] BU, T., LI, L., AND RAMJEE, R. Generalized proportional fair scheduling in third generation wireless data networks. In *INFOCOM* (2006).

[15] BUDDHIKOT, M., CHANDRANMENON, G., CH, G., HAN, S., LEE, Y. W., MILLER, S., AND SALGARELLI, L. Integration of 802.11 and third-generation wireless data networks. In *in Proc. of IEEE INFOCOM* (2003), pp. 503–512.

[16] BUDDHIKOT, M., KOLODZY, P., MILLER, S., RYAN, K., AND EVANS, J. DIMSUMnet: New directions in wireless networking using coordinated dynamic spectrum. *IEEE WoWMoM* (2005).

[17] BUDDHIKOT, M., AND RYAN, K. Spectrum management in coordinated dynamic spectrum access based cellular networks. *IEEE DySPAN 2005* (Nov. 2005), 299–307.

[18] CLEARY, F., FIEDLER, M., RIDEL, L., AND CIHAT TOKER, A. Perimeter: Privacy-preserving contract-less, user centric, seamless roaming for always best connected future internet. In *Wireless World Research Forum (WWRF) Meeting.* (May 2009).

[19] DAI, Z., FRACCHIA, R., GOSTEAU, J., PELLATI, P., AND VIVIER, G. Vertical Handover Criteria and Algorithm in IEEE802.11 and 802.16 Hybrid Networks. In *IEEE ICC* (2008).

[20] DE LA OLIVA, A., BANCHS, A., SOTO, I., MELIA, T., AND VIDAL, A. An overview of ieee 802.21: media-independent handover services. *Wireless Communications, IEEE 15*, 4 (2008), 96 –103.

[21] DUTTA, A., FAMOLARI, D., DAS, S., OHBA, Y., FAJARDO, V., TANIUCHI, K., LOPEZ, R., AND SCHULZRINNE, H. Media-independent pre-authentication supporting secure interdomain handover optimization. *Wireless Communications, IEEE 15*, 2 (2008).

[22] DUTTA, A., ZHANG, T., OHBA, Y., TANIUCHI, K., AND SCHULZRINNE, H. Mpa assisted optimized proactive handoff scheme. In *Mobile and Ubiquitous Systems: Networking and Services, 2005. MobiQuitous 2005. The Second Annual International Conference on* (2005), pp. 155 – 165.

[23] FLEISCHER, L., GOEMANS, M., MIRROKNI, V., AND SVIRIDENKO, M. Tight approximation algorithms for maximum general assignment problems. In *ACM SODA* (2006).

[24] FODOR, G., ERIKSSON, A., AND TUORINIEMI, A. Providing quality of service in always best connected networks. *IEEE Communications Magazine 41*, 7 (2003), 154–163.

[25] GUSTAFSSON, E., AND JONSSON, A. Always best connected. *Wireless Communications, IEEE 10*, 1 (2003), 49 – 55.

[26] KAMAKARIS, T., BUDDHIKOT, M., AND IYER, R. A case for coordinated dynamic spectrum access in cellular networks. *IEEE DySPAN 2005* (2005).

[27] KAUFFMANN, B., BACCELLI, F., CHAINTREAU, A., MHATRE, V., PAPAGIANNAKI, K., AND DIOT, C. Self organization of interfering 802.11 wireless access networks. In *IEEE INFOCOM* (2007).

[28] KELLY, F. P., MAULLO, A. K., AND TAN, D. K. H. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society* (1998).

[29] LAMPROPOULOS, G., SALKINTZIS, A., AND PASSAS, N. Media-independent handover for seamless service provision in heterogeneous networks. *Communications Magazine, IEEE 46*, 1 (2008), 64 –71.

[30] LI, L., PAL, M., AND YANG, Y. Proportional fairness in multi-rate wireless LANs. In *IEEE Infocom* (2008).

[31] MISHRA, A., SHRIVASTAVA, V., AGARWAL, D., BANERJEE, S., AND GANGULY, S. Distributed channel management in uncoordinated wireless environments. In *ACM Mobicom* (2006).

[32] MUSSABBIR, Q. B., AND YAO, W. Optimized fmipv6 handover using ieee802.21 mih services. In *Proceedings of first ACM/IEEE international workshop on Mobility in the evolving internet architecture* (New York, NY, USA, 2006), MobiArch '06, ACM, pp. 43–48.

[33] PAREKH, A. K., AND GALLAGHER, R. G. A generalized processor sharing approach to flow control in integrated services networks: the single-node case. In *ACM/IEEE Transactions on Networking* (1993).

[34] R. AGRAWAL, A. BEDEKAR, R. L., AND SUBRAMANIAN, V. A class and channel-condition based weighted proportionally fair scheduler. In *ITC* (2001).

[35] ROUGHGARDEN, T. Potential functions and the inefficiency of equilibria (survey). In *International Congress of Mathematicians* (2006).

[36] SESIA, S., TOUFIK, I., AND BAKER, M. *LTE, The UMTS Long Term Evolution: From Theory to Practice.* Wiley, 2009.

[37] WANG, W., LIU, X., VICENTE, J., AND MOHAPATRA, P. Integration Gain of Heterogeneous WiFi/WiMAX Networks. *IEEE Transactions on Mobile Computing* (Dec 2010).

[38] WOLFF, R. W. *Stochastic Modelling and the Theory of Queues.* Prentic Hall, 1989.

[39] WU, Q., AND ESTEVES, E. The CDMA2000 high rate packet data system. *Chapter 4 of Advances in 3G Enhanced Technologies for Wireless Communications* (March 2002).