A Study of Handover Strategies for Mobile Multiaccess Ambient Networks

Daniel Hollos, Petteri Poyhonen, Ove Strandberg, Ramón Agüero, Kostas Pentikousis, Oliver Blume

Abstract—In a network environment where multi-radio access becomes the norm, choosing the most appropriate network interface and base station becomes essential for distributing radio resources optimally and fairly, and for allowing each node to take full advantage of its multiaccess capability. This paper discusses the issues arising when mobile nodes are equipped with three different radio access technologies and have contradicting goals to the network provider in a high density scenario. We simulated three different decision strategies aimed to resolve such contradicting goals and compared those results to the optimal solution attained from a linear programming model.

Index Terms—Multiradio access; access selection; handover strategy; simulation study; linear program

I. INTRODUCTION

The proliferation of different radio access technologies (RAT) in recent years accompanied by advances in miniaturization and industrial design allowed vendors to market devices with several network interfaces. For example, even a few years ago tri-band mobile phones were far from mainstream; nowadays quad-band phones with personal area network (PAN) interfaces are common, and will soon be equipped with wireless LAN interfaces. In such multiaccess environments connecting a group of devices in an efficient manner is still an open research topic (see also Section II).

Decision automatically which interface to switch on, whether and when to use more than one interface simultaneously, and which network(s) to attach to are not easy problems when considering all parameters involved: Radio resource sharing, multioperator environment, security and AAA, end-to-end path optimality, energy efficiency, to name a few. Making “good” decisions involves several constraints and has to meet several objectives. Failing to satisfy the constraints causes service interruptions for the mobile node (MN); not meeting the objectives leads to, for example, inefficient use of MN battery power and network resources.

The state of the art presently lets MNs choose among different RATs based on rudimentary and static user and operator preferences. For example, for 3G/UMTS devices, the MN always opts for 3G rather than EDGE or GPRS. Although such connectivity priority settings are common, they may often lead to suboptimal infrastructure utilization.

The Ambient Networks (AN) project [1] has been developing a system architecture that considers multiple RATs during access decisions and can apply the appropriate handover (HO) protocol allowing for session continuity. In order to manage the system complexity, the control functions of an AN are split into dedicated functional entities (FE). Among these are the Multi Radio Resource Management FE (MRRM) [2], the Handover and Locater Management FE (HOLM) [3], the Policy FE [4], and the Flow Management FE. HOLM features a mobility tool box [3] comprising protocols, such as, Mobile IP [5] and Host Identity Protocol (HIP) [6]. MRRM manages different RATs and coordinates radio admission control, radio link setup and signal strength measurements.

In this work split several challenges arise, first of all that the FEs involved in decision making can be distributed between several nodes. It is safe to assume that the availability of constraints and objectives, in many cases, will depend on the physical location of each FE. For example, the operator does not want to share certain network-side constraints and objectives with the users (due to security, performance, and other limitations) and vice versa. Moreover, most FEs are not designed to have to deal with the detailed functional data of other FEs, and may operate at different timescales. For instance, while policy rules are rather static, radio signal strength often changes on microsecond scale. However, the optimal decision is influenced by most of the information in the system. This paper deals with decision methods and their performance where the decision is distributed among FEs having only partial information.

The rest of this paper is organized as follows. Section II presents an illustrative use case that establishes the requirements that the proposed solution should fulfill. Section III briefly surveys related work, Section IV introduces the system model used, while Section V presents the main results from our simulation-based evaluation. Finally, Section VI concludes this paper, outlining central items in our future work agenda.

II. USE CASE

The central goal of the handover access selection mechanism is to maximize overall network utilization and allow each MN to remain “best connected” at all times. For example, consider the case where a MN is receiving a streaming...
video fed over the 3G/UMTS cellular network, and during the playout MRRM detects two WLAN accesses, call them 802.11a, and 802.11b, as shown in Fig 1. Let us assume that in this use case there are these decision objectives, as follows.

1) With respect to radio capacity and quality, as measured at the terminal, the currently used cellular access offers the lowest bandwidth of all three options. Further, 802.11b seems slightly better than 802.11a (lower noise level), thus the choice would be 802.11a

2) With respect to the user preferences for the streaming application which prescribes the reduction of the number of HOs as much as possible, the current 3G connection should be further maintained

3) With respect to the global network resource utilization, since 802.11b is already operating at a high traffic load, the access choice would be to use 802.11a, in order to maintain connectivity and shift load from the low bandwidth cellular network to the broadband WLAN.

As this scenario illustrates, we face the problem of creating a decision framework where (i) different objectives lead to different decisions; (ii) the information has to be kept distributed among the deciding participants (FEs) (iii) the available information depends on the location of FEs and, thus, no FE can maintain a complete view; (iv) information from different FEs changes at different time scales. The Ambient Network approach in [7] solves this problem by assigning access selection to MRRM which continuously monitors the set of detected accesses (DS) and requests normalized weight values on this set from other FEs that show how strong that particular FE priorizes that particular access; this avoids disclosing internal parameters. Meeting constraints like policies, end-to-end QoS, and the required mobility protocols, MRRM [2] stepwise reduces the DS into subsets called validated set (VS) and candidate set (CS). Due to the limited knowledge, the sets at network side (CSNetwork) and at the terminal side (CSTerminal) may differ. MRRM finally selects one or more accesses by multiplying the weights and getting the accesses with the highest product, called the active set (AS) used for the communication session. Complete changes of the AS always result in a handover event.

This paper clearly cannot cover all topics mentioned above. Hence, we focus mainly on multi-RAT, cell-based, overlapping, single-operator scenarios and explore the benefits of different strategies. In particular we address the question how valid access decision can be taken in such a distributed and heterogeneous architecture and what benefit can be achieved compared to the current state of the art.

III. BACKGROUND AND RELATED WORK

There are quite a few recent works that have been looking at heterogeneous wireless architectures, evaluating different metrics and parameters. For example, [8] shows the impact of signaling timing issues over network-controlled handovers. The authors combine L2 and L3 information to improve handover execution procedures between 3G and WiFi networks. [9], [10] tackle the same type of heterogenous scenarios (mainly 3G and WiFi), however, they analyze only pure mobility solutions on top of them without trying to optimize the way handovers are performed.

The author in [12] developed a metric reflecting the efficiency of the wireless transmission of each terminal in wlan cells; this metric can be used as an input for access selection decisions. In order to support optimal access selection decisions, the IEEE 802.21 group specifies the Media Independent Handover (MIH) layer [11] which will allow gathering cross-layer information from heterogeneous technologies on a uniform way; additionally, the IETF’s MIPSHOP group is also looking at this type of heterogenous scenarios, developing end to end signalling schemes.

In this paper we approach the problem from a different perspective, analyzing how the combination of different decision strategies affects the performance in terms of mobile node connectivity. Our model is built according to the Ambient Networks reference architecture; [7] describes how this architecture supports MRRM realizing constraint-based access selection, going beyond legacy single-hop link quality metrics.

IV. SYSTEM MODEL

We consider a cell-based system with three heterogeneous radio technologies available. The MNs are moving around supporting all available RAT’s, however, they can connect to only one base station (BS) at a time.

A. Definitions and Working Assumptions

We use the following metrics and parameters in this study.

Multi-RAT MNs: Each MN supports three RATs, however, it can maintain only one active RAT at a time; switching to another RAT causes a short service interruption, thus the objective of the users is to reduce the number of HOs as much as possible.

Cell-based service: The operator provides cell-based service where each cell has a BS equipped with one of the three RATs supported by the MNs. See Section V-A for mode details.

Cell load: The cell load is calculated by measuring the traffic generated by all active MNs in a cell.

Traffic classes: The MNs can request two traffic classes, requiring 1TU and 1.5TU, respectively, where the Traffic Unit (TU) is an abstract measure for traffic load. The capacity of each BS is also defined in terms of TUs.
Radio model: Each entity has radial coverage area; the signal quality–thus the priority weight of the access from the channel’s perspective–is \( w = \max \left[ 0, 1 - \left( \frac{d_{bs, mn}}{r_{bs}} \right) \right] \), where \( d_{bs, mn} \) is the distance between the MN and the BS and \( r_{bs} \) is the radius of the BS.

Optimization metrics and goals: The goal is to select one of the available cells for each MN at any given timepoint so that the three optimization objectives listed in Section II are taken into account. Since the decision entities have only limited information, a merged utility function cannot be applied as objective function. Rather than that, how the objectives are combined is the peculiarity of the decision strategies detailed below.

B. Decision Strategies and Constraints

We model three distinct strategies: Terminal-centric, network-centric and legacy. We study the classical case of contradicting objectives between the end user (represented by the “terminal”) and the network operator (represented by the “network”). The terminal prefers stronger signals and tries to minimize the number of HOs performed. The network prefers accesses selection so that the load is balanced among the BSs and RATs; i.e., lower loaded access are preferred over high loaded ones. The strategies modelled are the following.

The Terminal-centric Strategy: The MN initiates the process by constructing \( CS_{terminal} \) based on the DS and applying its priority weights to each available access based on its constraints and objectives. The MN then delivers it to the network which checks each access in the received \( CS_{terminal} \) to see whether there is enough capacity for a new terminal, and if not, the examined access is omitted in any further constraint processing. Once the received \( CS_{terminal} \) has been checked, the network assigns its own priorities to the list and constructs the \( CS_{network} \) which it sends back to the MN. As a last step, the MN makes the final selection based on the access that are part of both \( CS_{terminal} \) and \( CS_{network} \) by merging the priority lists according to their rank orders.

The Network-centric Strategy: The MN first constructs the DS and delivers it to network, which then checks each access in the received DS to see whether there is enough capacity for a new terminal, and if not, the access is omitted in further constraint processing. Once the received DS has been checked, the network assigns its own priorities to the list and constructs the \( CS_{network} \) which it sends back to the MN. Then the MN constructs \( CS_{terminal} \) based on \( CS_{network} \) by assigning its priority weights to each available access based on its own constraints and objectives, and then delivers it to network. In the last step, the network makes the final selection based on the accesses that are part of both \( CS_{network} \) and \( CS_{terminal} \) by merging the priority lists according to their rank orders.

Legacy Strategy: The MN constructs the \( CS_{terminal} \) based on its DS and delivers it to network. The network constructs \( CS_{network} \) based on the received \( CS_{terminal} \) by applying its priority weights and makes the decision by the same multiplication procedure as above. The network informs the MN about the decision that updates its AS.

V. PERFORMANCE EVALUATION

We simulated the three algorithms described above using MATLAB. The evaluation in the MATLAB tool simulates the terminal attachment in a way that for every time slot all parameters for all MNs are re-evaluated. Further, we were interested to study the maximum achievable performance which is independent from the particular algorithms and limitations of information availability. We used a linear programming technique to calculate the optimal selections of accesses for the terminals. Linear programming techniques are often used in the area of network planning to get higher bounds on the performance [13], [14]. The maximum achievable performance shows us the potential of further developing our distributed algorithmic solutions. We used the following Mixed Integer Program (MIP) for this purpose.

\[
\begin{align*}
\max & \quad \sum_{i,b,r} R_{i,b,r} \cdot A_{i,b,r} \\
\text{s.t.} & \quad \sum_{b,r} A_{i,b,r} \leq 1 \quad \forall i \in \text{MN} \\
& \quad \sum_{i,r} R_{i,b,r} \cdot A_{i,b,r} \leq C_b \quad \forall b \in \text{BS}
\end{align*}
\]

where the decision variable \( A_{i,b,r} := \{0,1\} \) is set to one if \( MN_i \) is connected to \( BS_b \) using traffic type \( r \). Similarly, the parameter matrix \( R_{i,b,r} := \{1,1.5\} \) includes the load generated by \( MN_i \) if connected to \( BS_b \) using traffic type \( r \) in terms of TUs; the parameter matrix \( R \) lists only those \((i,b,r)\) combinations that are feasible, i.e., the current DS of the MN. Finally, \( C_b \) is the maximal capacity of the base station \( b \).

The objective function (1) aims at maximizing the load the MNs generate; constraint (2) ensures that a MN is connected to at most one BS at a time, and constraint (3) prevents BSs from exceeding their capacity limits. This model can be solved by any MIP solver, like scip or cplex [15]. We used the latter one for getting the optimal access configurations for every 1200 simulated timepoints, individually. The results are reported in Figs. 3-6 as “optimal” case.

A. Simulation Parameters

There are 300 MNs moving according to a mobility model similar to the random waypoint model with the following parameters: The maximum speed is 10m/sec with a random \( \pm 90^\circ \) direction change probability, however, unlike in the random waypoint model, the MNs do not have idle moments. Every algorithm evaluation is executed every 0.1s, and the simulated time was 120s.

We consider a square with dimensions 1000 m \( \times \) 1000 m, served by three different RATs, which we refer to as RAT-1, RAT-2, and RAT-3. The following cells were serving the MNs traffic demands: 30 cells of type 802.11a (RAT-1), 4 cells of 3G (RAT-2), and 60 cells of type 802.11b (RAT-3). The simulation scenario is shown in Fig. 2. In the figure RAT-1 coverage is indicated by solid red circles, RAT-2 by blue circles, and RAT-3 with black circles. Note that all points of the entire 1km\(^2\) area are covered by at least one of the RATs, and there are subareas with extensive overlapping.
The load generated by the MNs changes randomly during the simulation. Recall that each MN may request traffic type of 1TU or 1.5TU. A cell accepts the connection request if it is within the cell capacity limit defined as follows. The 802.11a and 802.11b cells have capacity limit of 5TU, while the 3G cells have the capacity of 30TU. In the case a MN requested the 1.5TU traffic type but the cell is only able to admit 1TU, the MN can connect using the lower connection capacity rate.

B. Simulation Results

Since we were interested in the stability and transients generated by our decision methods—not faded by statistical methods—, we performed one long simulation (120s) for each access selection method, instead of repeating shorter ones several times. We logged the state of the network every 0.1s, the results are reported in Figs. 3-7.

Figs. 3-5 show the performance of the legacy, terminal-centric, and network-centric decision methods for the different RAT types, including also the optimal allocations. The performance of the decision methods, in terms of connectivity, mainly differs for RAT-1: The connections of RAT-1 is clearly increasing from legacy- to terminal-centric method and further increasing for the network-centric one. The increasing amount of information at the deciding entity helps to utilize the RAT-1 resources which results in the increased load; RAT-1 is at most loaded in the network-centric case, since the load balancing criteria directs more MNs to RAT-1 attachment.

An important goal for the network providers is to balance the usage of their network resources (i.e. the load should be evenly distributed among the BSs) while minimizing the fluctuations in each BS: Too big fluctuations can cause unpredictable overloads, and thus service interruptions. Examining Figs. 3-5 from this perspective, RAT-1 and RAT-3 are experiencing differences with two opposite trends. On one hand, the legacy modus uses RAT-3 at most unbalanced; it is better balanced for the terminal-centric method, and is the most balanced (i.e. flat) by the network-centric method. On the other hand, the trend is opposite for RAT-1. It is at most balanced in legacy modus, followed by the terminal- and network-centric methods, see also Fig. 5: The network-centric method drives as many MNs as possible towards using the bigger RAT-3 cells; only those MNs are rejected from RAT-3 will select RAT-1. Thus, if a bunch of disconnected MNs, accidentally, approach the relatively small area served by RAT-1 cells (red circles in Fig. 2), they will appear as “new load” for those cells.

The overall performance of the different distributed algorithms are shown in Figure 6 where the network-centric method clearly shows network load performance superiority, followed by terminal-centric method, leaving the legacy only showing a load level is on average 83% of maximum 300 fully connected MNs. The approximate load levels are on average 88% and 95% for the terminal- and network-centric strategies respectively, and the optimal solution is at 99.1%.

Fig. 7 illustrates that both network-centric and terminal-centric strategies lead to more HOs than the legacy modus. The network-centric approach clearly shows network load performance superiority, followed by terminal-centric method, leaving the legacy only showing a load level is on average 83% of maximum 300 fully connected MNs. The approximate load levels are on average 88% and 95% for the terminal- and network-centric strategies respectively, and the optimal solution is at 99.1%.
the network-centric approach results in better connectivity and balanced usage of the resources, it is at the price of three times more HOs which maybe undesirable for the user application. Some cooperation with the MNs could alleviate this effect, e.g., by applying at least partially merged objective functions which would then also reflect the HO minimization goal of the user application.

### VI. Conclusion and Future Work

The different strategies have all their own advantages and drawbacks. Clearly, an overall load efficiency objective is most benefitting from the network-centric strategy, while the combination of low amount of HOs with good connectivity times is best reached by the terminal centric strategy.

In addition, we showed that the network-centric strategy is likely to be able to support more connected MNs than the terminal-centric one, because its decisions are based on global knowledge, thus able to shift the load towards the small-coverage cells in order to relieve the ones with more MN requests. It is also interesting to note that the loads on the other two RATs (RAT-2 and RAT-3) are quite similar for the two strategies, probably due to the fact that they both cover quite a large percentage of the simulated area.

As for the future work, adding multiple, competitive operators would most likely influence our results significantly. Keeping the decisions near to the optimal case, similar to the single operator case reported in this paper, would probably require the dynamic adaptive selection of decision algorithms themselves; this selection would probably also depend on the MN requirements and policies, at least for the case the MN has BSs from different operators in its DS.

### References


