Toward an Autonomic Control of Wireless Access Networks

N. Blefari-Melazzi
D.I.E. - University of Rome "Tor Vergata", Italy
blefari@uniroma2.it

D. Di Sorte, M. Femminella, G. Reali
D.I.E.I. - University of Perugia, Italy
{dario.disorte,mauro.femminella,gianluca.reali}@diei.unipg.it

Abstract—In this paper, we present a solution to control automatically a wireless access network and we focus on load balancing target. The reference scenario is that of a single administrative domain that offers a heterogeneous wireless access. We propose a monitoring procedure able to track the access network context, i.e., the current state of resources, the amount of service demand, and the wireless access points available to mobile terminals. To this end, we exploit a cooperative brokerage framework, distributed between terminals and network. The access network context is the input of a selection process that aims to distribute suitably mobile users among available access points. The objective is to balance the users demand, thus improving the quality of the service. The numerical analysis shows the effectiveness of the proposed system in an 802.11b access network.

Keywords— access network context, access point selection, load balancing, 802.11b

I. INTRODUCTION

This work has been carried out in the framework of the EU Simplicity project [1]. Simplicity stands for Secure, Internet-able, Mobile Platforms Leading Citizens Towards simplicity.

The Simplicity project started from a vision: users today employ a variety of different terminals and devices to access a range of different services in the home, in buildings or in public spaces. Some services can be very complex and may require location awareness, QoS support, message exchanges with network databases, structured interaction with remote networking devices. Heterogeneous services, terminals, and networks create a complexity barrier not only to end-users but also to operators, who have to devise and deploy tools and procedures to engineer their network through load balancing, intelligent resource exploitation, congestion avoidance, fairness, fault tolerance and so on. The emergence of new research areas, such as pervasive computing, will further increase the diversity of the devices and services with which users and operators have to deal with.

In this framework, a key attribute of Simplicity [1][2] is reconfigurability, at various levels. To integrate different paradigms from the user point of view, it is necessary to break logical wires that still tie mobile users to networks and services, also at upper layers. This way, heterogeneous and mobile access networks can be really integrated, as IP has glued heterogeneous networks. To enable this full-spectrum reconfigurability, our system encompasses three components: the Simplicity Device (SD), the Terminal Broker (TB) and the Network Broker (NB).

The SD is related to the portability of user identities among different classes of terminals. Each user will be characterized by a personalized profile, providing access to different services and networks. The user profile will be stored in the SD. Users could personalize terminals and services by the simple act of plugging the SD (e.g., a Java card or a USB stick) into the chosen terminal. This allows users to enjoy the automatic selection of services appropriate to specific locations, the automatic adaptation of information to specific terminal devices and user preferences, and the easy exploitation of different telecommunications paradigms and services.

The TB is the entity that manages the interactions between the information stored in the SD and the terminal. The TB enables the SD to perform actions like terminal capability discovery, adaptation to networking capabilities and to the ambient, resource and service discovery and usage, adaptation of services to terminal features and capabilities.

The NB has the goal of providing support for service description, advertisement and discovery. Moreover, it orchestrates service operation among distributed networked objects, taking into account the issues related to the simultaneous access of several users to the same resources, services, and locations. It also shares/allocates available resources, and manages value-added networking functionality, such as service level differentiation and quality of service, location-context awareness, and mobility support.

The capabilities of the Simplicity broker not only allows users to use ICT systems spontaneously and simply, but also provide operators with new options to define new paradigms for distributed, technology independent, self-organizing, and autonomic networking systems, able to react locally and autonomously to context changes [13]. This should strengthen the operator interest in deploying a Simplicity system, since it would help network resource management.

In this paper, we focus on load balancing in wireless networks, which is an emerging research field (e.g., see [4][5][6][7][8][9]). The reference scenario is that of a single administrative domain that offers a heterogeneous wireless access. In principle, mobile terminals may be provided with a number of wireless interfaces (multi-mode terminals).

The novelty of our solution is the exploitation of the distributed Simplicity broker. We propose a monitoring procedure able to track the access network context, i.e., the current state of resources, the amount of service demand, and the wireless access points available to terminals. We assume that the TB is able to perform frequency scanning and to listen...
to the L2 beacons transmitted by surrounding wireless access points periodically, and to learn their L2 IDs (in accordance with [5]). These Access Points (APs) are candidates to hand over to. Then, an appropriate selection procedure running in NB uses this set of information to drive MNs towards the most appropriate AP. Also, information retrievable from SDs can be useful to predict users’ behavior (e.g., user demand) and to differentiate the network service (e.g., according to the tariff profile and the user role).

In this paper, we show by simulations the effectiveness of the proposed approach in terms of load balancing capabilities in an 802.11b access network. We also provide directions to dimension a number of system parameters.

The paper is organized as follows. In the next section, we illustrate the access network control procedure. Section III illustrates numerical results. Finally, section IV reports some concluding remarks and considerations on future work.

II. ACCESS NETWORK CONTROL

The typical goal of a network manager is to optimize network performance in terms of QoS (users' side), throughput and load balancing (operator's side), pre-empting critical situations and minimizing human actions. In general, network control actions will depend on users' side information, on the spatial distribution of users over the area covered by the network, and on the characteristics of the network, such as network topology, network resources, and available tuning capabilities. The goal is to get proactively the highest possible amount of data available as input to the decision engine. In this regard, we suggest that these goals could be achieved more easily by exploiting specific features of Simplicity.

In this paper, we especially deal with load balancing in the wireless access section of the network.

Some papers present mechanisms to manage layer 2 handovers in a homogeneous 802.11 access section. In this regard, the task group 802.11k is defining a radio resource measurement enhancements to provide interfaces to higher layers for radio and network measurements. These measurements are required to provide management services, such as load balancing and handover management.

In [4], the Authors propose a distributed architecture based on agents running on 802.11 APs, which exchange traffic load information to cooperatively balance the traffic by forcing the handover of a subset of MNs staying under the coverage of an overloaded AP. The drawback of this approach is that the procedure is not aware of which APs are available to an MN, and this implies that an MN forced to de-associate from the current AP could be denied the network service. In our opinion, if load balancing decisions are taken on the network side, information about the access network context from the terminal side are necessary to make the approach effective. Otherwise, as proposed in [6], users could be explicitly requested to cooperate by physically moving towards specific locations within the network for load balancing purposes.

If decisions are taken by MNs, in [8] Authors propose to embed load balancing information in 802.11 beacon frames. MNs should then use this information in addition to the signal strength to select the AP. This solution also leads to a lack of interoperability between different vendors. In addition, if an operator wishes to hide network-related information for security and/or business reasons, this solution could be critical.

We stress that our aim is to present a technology-independent procedure, the scope of which is an IP network with a number of Access Routers (ARs) that control a set of heterogeneous APs. Handovers can be both inter and intra-technology, intra and inter AR, at both layer 2 and layer 3.

We exploit the capabilities of SD-enabled terminals. MNs are assumed to perform frequency scanning and listen to the L2 beacons transmitted by surrounding APs periodically, and to learn their L2 IDs (identifier of an AP that uniquely identifies that AP, [5]); these APs are candidates to hand over to. Note that the capability of performing frequency scanning is actually a minimum requirement for all wireless technologies.

In the following, we first give details about the procedure and functions for the execution of load balancing actions, and then present some architectural considerations about the NB.

The reference scenario is that of a single administrative domain (e.g., a campus network) which offers a heterogeneous wireless access (e.g., 802.11b, GPRS/UMTS, Bluetooth). We assume that the wireless coverage offers many options for selecting a wireless access (dense wireless coverage), so that a load balancing mechanism makes sense. In principle, a mobile terminal may be provided with a number of wireless interfaces (multi-mode terminals).

A. The Simplicity-enabled mechanism

The main peculiarities of the Simplicity brokerage framework are:
- the capabilities of SD-enabled terminals, which can be exploited to assist the network in monitoring the availability of wireless coverage;
- user information retrievable from SDs (profiles and preferences), which can be useful to predict users' behavior and to differentiate the network service according, for instance, to users' role and tariff profile.

We assume that the network operator is in charge of taking network configuration and traffic engineering decisions. Users may only provide inputs (through the SDs) to influence the management actions of the operator. Thus, the NB is the entity in charge of acting as Policy Decision Point (PDP), i.e., it is the functional entity in charge of taking decisions.

In turn, the TB is in charge to assist the NB, by providing it with inputs regarding the access network context (i.e., the radio access technologies currently perceived by the terminal through a frequency scanning action), and by acting as Policy Enforcement Point (PEP), e.g., to switch among APs.

In addition, a proper monitoring process has to be able to provide the NB with the information about the current status of
wireless resources at the APs of the network. To sum up, the inputs to the NB are:

- users' side information (e.g., (i) user role; (ii) set of subscribed service; (iii) willingness to pay). This set of information is sent once (e.g., when the user registers to Simplicity) and updated if needed;
- terminal capabilities (e.g., radio technologies supported);
- access network context (e.g., available wireless accesses, perceived power level, and status of network resources). This set of data is dynamic and has to be refreshed (either periodically or upon request);
- specific management policies, defined by the network operator, with the aim of improving network performance while maintaining user satisfaction.

The output of the NB consists of network control decisions. By applying properly designed algorithms, the input to the NB can be used to automatically optimize the distribution of mobile users within the wireless section, but also to activate/deactivate APs, to tune APs power, and so on. In this paper, we focus on load balancing actions; mobile terminals are driven towards the most appropriate AP, according to specific policies of the operator. The selection process may be invoked: (i) when the terminal is turned on; (ii) periodically, due to specific operator's policies; (iii) when a handover is needed. The overall process is sketched in Fig. 1. It is worth noting that when the terminal turns on, the TB is allowed to exchange information with the NB through a default network connection. Then, the NB drives the terminal towards the most appropriate wireless access.

As a final note, we stress that the proposed approach is general and can be used in a heterogeneous wireless access network environment. In the following, we will refer to a homogeneous 802.11b scenario.

### B. The decision metric

For the AP selection in the Simplicity enhanced system we use the criterion of the less loaded AP among the set of candidates for a given MN. Assuming that MNs have homogeneous profile and service demand, we define the load of $i$-th AP as the number, $N_{mn}(i)$, of MNs attached to it. Thus, we do not consider a specific traffic model, and assume that the wireless bandwidth of an AP is fairly shared among users. The NB assigns to a candidate AP a cost equal to:

$$M_{AP}(i) = N_{mn}(i) + xH_{MN}$$

where $x=0$ for the current AP, and $x=1$ for the candidates.

The parameter $H_{MN}$ (hysteresis) is introduced to avoid annoying ping-pong effects (i.e., continuous switches among overlapping APs). The higher the value of $H_{MN}$, the lower the load balancing, and the higher the stability of the process. Thus, the choice of the value of $H_{MN}$ has to be made considering a trade-off between performance and stability.

Note that, in the case analyzed, the NB is able to maintain the association between each AP and the related MNs served. In other words, it is not necessary to probe APs to get such an information. Thus, no special functions within APs are required. If more sophisticated selection criteria based also on bandwidth measurements at APs are used, then the probing phase between NB and APs becomes mandatory.

The performance figure we consider is the gain obtained using the Simplicity system with respect to the legacy one:

$$Gain = \frac{BW_{\text{min-SIMPLICITY}}}{BW_{\text{min-LEGACY}}}$$

where $BW_{\text{min}} = \min_i[BW_{\text{net}}/N_{mn}(i)]$ is the minimum value of net bandwidth available, on average, for each of the users under the coverage of the most crowded AP in the network.

Clearly, the better the load balancing, the higher the service level (i.e., the amount of bandwidth) perceived by users.

### C. The NB architecture

In principle, the NB may be either centralized over a specific network entity or distributed among different network entities, each one within a single sub-net and controlling the traffic of that sub-net. In this latter case, we assume that the NB instances are co-located with ARs, so that each AR is in charge of managing only those mobile users under its control.

In our opinion if the network size is small the deployment of a single centralized NB is a reasonable choice.

On the other side, if the network is large, the distributed solution is mandatory for scalability reasons. In this case, each AR has to be able to retrieve the network status of both its APs and the APs belonging to neighboring ARs, whose coverage area overlap with its own. The configuration of the surrounding wireless coverage map at each AR (discovery phase) may be either manual or automatic; in this latter case, the network has to be able to self-discover its wireless coverage. In operation, neighboring ARs have to exchange the service capabilities of their APs (steady phase).

A deep analysis of both self-learning discovery phase and steady phase, along with a quantitative evaluation of discovery time and signaling burden, is beyond the scope of this paper (refer to [10]). In this regard, we only remark that the overall signaling burden associated with both discovery and steady phase in the whole network is definitely low (few Kbps). Thus, in our opinion the benefits that can be obtained by implementing a Simplicity-enabled load balancing mechanism justify the relevant cost in terms of the network resource consumption and the complexity of implementation.
III. PERFORMANCE EVALUATION

In this section, we present the numerical results from a simulation campaign, which show the load balancing capabilities of the Simplicity-enabled network management system. To this end, we have developed a proprietary C++ simulator exploiting the NS-2 event scheduler.

A. Simulation scenario

The simulation topology used at this stage is a square area with side 150 meters (e.g., a portion of a campus area). We have implemented the NB distributed among 9 ARs. The number, \( N_{\text{mn}} \), of simulated MNs is equal to 800. For what concerns the adopted radio technology, we emulated 802.11b. The number of 802.11b APs is 21. They are distributed in order to assure uniform service coverage. Most of the area is covered by two different APs, and some parts by 3 APs (see Fig. 2). For what concerns the transmitted power, we used the value of 100mW, with receiver sensitivity equal to 3 APs (see Fig. 2). For what concerns the received power, we have used the value of 100mW, with receiver sensitivity equal to \( PW_{\text{min}} = 11\text{nW} \). This implies that, assuming a quadratic attenuation model (with transmission/reception gain equal to 1 at a frequency of 2.4GHz), the radius of the circular cell is equal to 30m. In addition, we have defined a further power level \( PW_{\text{opt}} = 19.5\text{nW} \), corresponding to a radius of 22.5m, i.e., 75% of the radius, which defines an alarm threshold. In the Simplicity mode, when a MN realizes such a power level on the current AP, it triggers the selection process at the NB, since it is next to the cell border. We implemented the AP coverage area, which comprises 5 attractors, whose co-ordinates are \((45,45), (45,65), (65,45), (55,55), \) and \((65,65)\). The latter is a small area, with a single attractor placed in \((125,125)\).

The simulation time is equal to 3000s. The MNs freely move during the interval \([0, 1000]\) s, then they stop for 500s \((0 \text{ to } 1500] \) s: phase 1). At simulation time 1500s, all the MNs restart moving, and a subset of them are attracted towards the first attraction area. This setting lasts for 180s, and then all MNs stop again for 500s \((1500 \text{ to } 2180] \) s: phase 2). At 2180s, all the MNs restart moving, and the same subset of MNs previously attracted by the first set of attractors move towards the point \((125,125)\). Finally, at 2360s, all MNs stop up to the simulation end \((2180 \text{ to } 3000] \) s: phase 3).

The rationale of this setting may be framed in a campus network. The movement towards the first set of attractors may be representative of a number of users (e.g., students and professors) gathering in laboratories or classrooms for lessons. For this reason, we placed more than a single attractor. Clearly, once arrived, it is necessary to emulate the stay in such a place. Then, the movement towards a single point may be representative of movement towards refectory.

The fact that MNs stop for a given amount of time allows testing the capability of the Simplicity system to reach (and maintain) a steady state condition. Clearly, phase 1 is the less challenging for load balancing, whereas phase 3 represents the most crucial situations, since (i) a subset of MNs move towards a single point of the simulation area, and (ii) these MNs move together starting from the same area. This implies that they follow nearly the same path contemporarily. We expect that this phase is more critical than phase 2, where MNs converge towards a larger area from different directions, thus sharing a potential large amount of wireless resources.

B. Numerical results

We compared two models:

- the legacy one, designed according to the typical implementation of real 802.11b systems, in which a terminal remains attached to the current AP until the received power level goes below \( PW_{\text{min}} \) (the receiver sensitivity). When the terminal realizes to be disconnected, it performs beacon scanning and, among the set of found APs, it selects the one with the highest signal strength;

- the Simplicity mechanism: when the power level of the current AP goes under \( PW_{\text{opt}} \), the terminal performs L2 beacon scanning, communicates such a result to the NB, and requests a driven handover; when the terminal turns on, it initially selects the AP with the strongest signal strength. In turn, the NB, based on context information from the TB and the measurement collected each \( T_{\text{tbw}} = 1 \text{s} \) on the APs, decides the best AP to attach. The terminal attaches to such an AP. The best AP is the less loaded among the set of candidates, according to the metric illustrated in subsection II.B. An AP is considered a candidate if its power level is above \( PW_{\text{min}} \). As mentioned above, the selection process is also triggered periodically with period, \( T_{\text{sel}} \), set equal to \( T_{\text{sel}} = 1 \text{min} \).

All measurements are averaged over 20 simulation runs. Fig. 3 reports the performance gain obtained with...
Simplicity-enabled load balancing with respect to a legacy system, with $H_{MN}$ as a parameter, when the percentage of attracted MNs is 60%. A remarkable feature is that the more critical the condition due to MNs behavior (from free to softly biased to strictly biased movement pattern), the higher the gain achieved by the Simplicity mechanism, up to values around 2.4. In other words, we verify a performance gain increasing with the simulation time (at least in the period when MNs are blocked and the mechanism can converge towards a steady condition). In addition, we expect that the higher the percentage of attracted MNs, the higher the gain. In order to show these effects and confirm our thesis, consider Fig. 4. It presents the performance gain vs. simulation time for different percentage of attracted MNs, with $H_{MN}=15$.

![Performance gain vs. simulation time with $H_{MN}$ as a parameter.](image)

In addition, let us consider the effect of the hysteresis parameter on performance and stability of the proposed procedure. As mentioned above, we expect that the higher the hysteresis, the higher the stability of the procedure and lower the performance gain. This behavior is clearly depicted in Fig. 3. It shows that a value of $H_{MN}=10$, even if performs better than $H_{MN}=20$, implies slight jittering and decreasing behavior of the performance gain in the critical phase 3. We have verified that when the percentage of attracted MNs increases and reaches 80% and 100%, a value of $H_{MN}=15$ is the best choice for the most critical phases (2 and 3).

![Performance Gain (Simplicity/Legacy) vs. Simulation time](image)

The curves are obtained assuming that the selection period is equal to $T_{SEL}=1$ min and that the information about the network load on APs is refreshed each $T_{BW}=1$ s. Let us now study the behavior of the system when $T_{SEL}$ and $T_{BW}$ vary.

To have a low computational burden in the NB and a low signaling exchange among NB and TBs, a network operator would like to select a high value of $T_{SEL}$. However, this would clearly lead to a performance loss since the mechanism is not able to quickly balance the load when the network conditions and MNs position vary. Thus, a trade off between promptness and workload is necessary in the selection of $T_{SEL}$.

Moreover, a high value of $T_{BW}$ would be preferable to maintain a low signaling burden in the fixed part of the network. On the other side, too high values of this parameter affects the system stability, since the load balancing mechanism is not able to follow the dynamics of the network status. Thus, a trade off between stability and signaling burden is necessary in the selection of $T_{BW}$.

![Performance gain vs. simulation time with the percentage of MN attracted as a parameter ($H_{MN}=15$).](image)

As regards $T_{SEL}$, the simulation analysis says that, for low values of $T_{SEL}$ (few seconds), the system is not stable in critical situations (phases 2 and 3), since too frequent load balancing actions are not able to reach a steady condition. However, we are interested in evaluating the possibility to increase the value of this parameter without loosing the performance gain with respect to the legacy system. In Fig. 5, we show the performance gain as a function of the selection period with $H_{MN}$ as a parameter; $T_{BW}$ is set to 1 s. We consider the following meaningful simulation times: (i) the end of the free movement of MNs ($t_s=1000$ s); the end of phase 1 ($t_s=1500$ s), when MNs stop after free movement; (iii) the end of phase 2 ($t_s=2180$ s), when MNs stop after a strictly-biased movement pattern; (iv) the end of phase 3 ($t_s=3000$ s), when MNs stop after a strictly-biased movement pattern. From inspection of curves, we can assert that for values of $T_{SEL}$ beyond 60 s, the performance loss of the system is not noticeable. This is a very good result, since it makes possible to reduce the signaling burden in the network among TBs and NBs. In addition, the higher the value of the selection period, the lower is the computational burden to be supported by the NB. Note that, when $H_{MN}=10$ and $T_{SEL}=120$ s, we have the best performance gain coupled with stability (due to less frequent decisions); obviously, this high value of $T_{SEL}$ also implies a longer time to reach the steady condition, i.e. lower responsiveness.

When $H_{MN}=10$, the system becomes unstable for values of $T_{BW}$ immediately higher than 1 s, whereas for $H_{MN}=15$ or 20 the system remains stable for values up to 10 s. To validate this statement, we report Fig. 6 which shows that a value of $T_{BW}$ equal to 10 s leads the system to an unstable behavior in phases 2 and 3 for low values of the hysteresis parameter; $T_{SEL}$ is set equal to 60 s. Increasing the $H_{MN}$ value compensates such a behavior, and for a value equal to 15, the system exhibits a gain comparable with the nominal one (i.e., $T_{BW}=1$s). This is an interesting result, since it implies that we can lower to some extent the bandwidth consumption due to signaling exchange.
This clearly leads to a slight performance loss with respect to the case with $H_{MN}=10$ and $T_{BW}=1$ s.

To sum up, in the specific network scenario analyzed, in our opinion a good choice of the system parameters that guarantees a good trade off among performance gain, responsiveness, stability, signaling and computational burden is: $T_{SEL}=60+120$ s, $H_{MN}=15$ and $T_{BW}=1+10$ s.

![Performance gain vs. $T_{SEL}$ with $H_{MN}$ as a parameter at simulation times](image)

Fig. 5 - Performance gain vs. $T_{SEL}$ with $H_{MN}$ as a parameter at simulation times (a) 1000s, (b) 1500s, (c) 2180s, and (d) 3000s; $T_{SEL}=1$s.

IV. CONCLUSION AND FUTURE WORK

We have presented a procedure to automatically control the wireless access network. To this end, network operators can use the enhanced peculiarities of SD-enabled terminals. We have especially focused on the load balancing target. The simulative analysis in an 802.11b scenario has shown that the effectiveness of the Simplicity mechanism is noticeable and increases when the network resources are more stressed.

In addition, we have provided directions for dimensioning of the system parameters considering both the performance and the costs of the procedure.

We are aware that this work represents a first-step analysis to evaluate the potentiality of the Simplicity approach. Another step of the analysis includes a further evaluation of the load balancing performance solution in terms of UDP/TCP throughput with specific traffic models (see [12] for an analysis with UDP traffic). In addition, we plan to account for new inputs from the SD. For instance, in a campus, different class of users (e.g., professors, students, and employees) may have different network service treatments.

Also, a demonstrator of the system is currently being produced in an 802.11b network to show its feasibility.

![Performance gain vs. simulation time with $T_{BW}=10$ s and $H_{MN}$ as a parameter](image)

Fig. 6 - Performance gain vs. simulation time with $T_{BW}=10$ s and $H_{MN}$ as a parameter; $T_{SEL}=60$s.

REFERENCES