Characterising University WLANs within Eduroam Context

Marangaze Munhepe Mulhanga, Solange Rito Lima, and Paulo Carvalho

University of Minho, Department of Informatics, 4710-057 Braga, Portugal
e-mail: solange@di.uminho.pt

Abstract. The *eduroam initiative* is assuming an ever growing relevance in providing a secure, worldwide roaming access within the university WLAN context. Although several studies have focused on educational WLAN traffic characterisation, the increasing variety of devices, mobility scenarios and user applications, motivate assessing the effective use of eduroam in order to sustain consistent network planning and deployment. Based on recent WLAN traffic traces collected at the University of Minho (Portugal) and University of Vigo (Spain), the present work contributes for identifying and characterising patterns of user behaviour regarding, for instance, the location and activity sector of users. The results of data analysis quantify the impact of network access location on the number of associated users, on the number and duration of sessions and corresponding traffic volumes. The results also illustrate to what extent users take advantage of mobility in the WLAN. Complementing the analysis on a monthly basis, a fine grain study of WLAN traffic is provided through the identification of users' behaviour and patterns in small timescales.

Key words: Network Traffic Characterisation; WLAN; Eduroam

1 Introduction

In recent years, there has been a significant growth of Wireless Local Area Networks (WLANs), with increasing influence on people's day life and productivity. The low to moderate cost of personal devices and the need for easy and ubiquitous access to information are factors that influence the growing use of wireless technology.

Network Access Points (APs) tend to be critical points in WLANs due to multiple aspects such as user mobility, traffic dynamics, location and density of devices, which impact on network performance. From the user perspective, WLAN are expected to provide seamless connectivity, supporting multiple types of applications and services, such as VoIP, video conferencing, Web Services, with distinct quality of service (QoS) requirements. Understanding traffic characteristics and the usage of network resources is an essential step to assure quality of service (QoS) and improve quality of experience (QoE) of end users when accessing network services.

In this context, this paper presents a traffic analysis and characterisation study involving WLANs of two European Universities - University of Minho (Portugal) and University of Vigo (Spain), within *eduroam (Education Roaming) initiative* [1]. The analysis was carried out based on real traffic traces gathered between April and June of 2010, which corresponds to a typical academic term. The study assesses several network usage metrics related to AP utilisation, session characterisation, access location influence, and user mobility patterns. This analysis, when compared with other case studies, aims to provide guidelines for planning future WLANs deployment.

This paper is organised as follows: related work reporting university WLAN studies is discussed in Section 2; the University campi and corresponding WLAN infrastructure are briefly described in Section 3; the process of data collection is presented in Section 4; the results of traffic analysis are discussed in Section 5; and the main conclusions are included in Section 6.

2 Related Work

Traffic analysis and characterisation has been the matter of relevant and extensive research over the years. However, within the University WLANs context, there are several works addressing this topic.

Tang and Baker [2], based on data collected during twelve weeks in one building of Stanford University, aimed to understand the behaviour of WLAN users, answering questions regarding the benefits of mobility, and the volume and characteristics of traffic involved. QoS metrics, such as delay and bandwidth were also measured.

Kotz and Essien study [3], reporting the analysis of data traffic collected during eleven weeks on the campus of Dartmouth University, complemented the study in [2], extending the analysis to all buildings on campus. Later on, the work reported in [4] states that the applications used over the already mature WLAN changed dramatically.

Balazinska and Castro [5] analysed a four week trace gathered at a corporate 802.11b WLAN, encompassing three buildings hosting computer science and electrical engineering research groups. The study focuses on population characteristics, load distribution across APs, user activity levels, and user mobility. This study found that: users' average transfer rates follow a power law; load is unevenly distributed across access points and is more influenced by which users are present than by the number of users; the APs location plays an important role in the aggregate load observed; and users spend a large fraction of their time at a single location.

The motivation for the study carried out by Schwab and Bunt [6] was to understand usage patterns in the University of Saskatchewan campus, comprising a small number of access points (18) strategically placed, in order to plan its expansion. For this purpose, usage data was collected over the period of one week in January 2003, recording address and protocol information for every packet sent and received on the wireless network. The trace was analysed to answer questions about where, when, how much, and for what the wireless network was being used.

Papadopouli et al [7] investigated roaming activity at aggregated level in the University of North Carolina infrastructure. Based on syslog data from three monitoring periods (between October 2004 and April 2005). The authors identify the regions with high roaming activity and derive topological models of the University infrastructure, involving 488 APs. The study also discusses the impact of the spatial and temporal growth of the wireless infrastructure, discussing the nonlinear correlation between the number of roaming events between two APs and their geographic distance.

Kumar et al [8] study classifies users into social groups and investigates the WLAN usage behaviour of these groups in the USC University campus (MobiLib). Based on a month long WLAN trace, the authors analyse aspects such as the differences on the average session duration for male and female users across the campus.

Kim and Helmy [9] study, based on traffic traces collected from Dartmouth University WLAN during 4 years (between 2001 to 2004 and 2005 to 2006), try to understand how changes in wireless devices and network affects WLAN users, and influence location prediction. The study highlights the drastic change in the number of APs and growing of mobile user community, and defines the number of distinct APs a user has visited as mobility metric.

The analysis of the above studies provides useful insights on WLANs usage patterns and important guidelines for further study on this research topic. However, the mentioned works, report results on distinct aspects of WLANs usage based on traffic traces collected before 2006. Facing the constant evolution of wireless devices, the increasing number of users, the variety of available applications and services, and human behaviour regarding wireless technologies, it is essential to keep an up-to-date analysis of today's university wireless networks usage, in particular, within the eduroam context. These aspects will provide useful feedback regarding WLAN planning and future deployments. The present study is a further step in this direction, taking as case study the analysis of the WLANs from University of Minho and University of Vigo using traffic traces collected between May and June of 2010.

3 Case Study: Eduroam at University of Minho

The University of Minho (UMinho), located in the north of Portugal, was founded in 1973 and started its academic activity in 1975. Currently, with a population of nearly 15,000 students, 1,200 teachers, and 600 technical and administrative staff, it is one of the biggest Portuguese universities. The academic and scientific activities at UMinho are developed in two campi: the campus of Gualtar in Braga and the campus of Azurém in Guimarães. The students' accommodation buildings have capacity to accommodate 1400 students, around 60% in Braga and 40% in Guimarães.

Both campi participate in the eduroam initiative, which provides wireless network access for research and education to the university population and visitors. The WLAN infrastructure comprises a total of 429 APs, 310 located in Braga and 119 in Guimarães. The WLAN technology used in both campi is based on the standards IEEE 802.11b and 802.11g and, more recently, 802.11n. 802.11n is deployed in strategic locations, such as libraries, due to its advantages both in terms of transmission data rates and spatial coverage [10, 11]. IP addresses are assigned to wireless devices via DHCP. Authentication is performed by a Radius server.

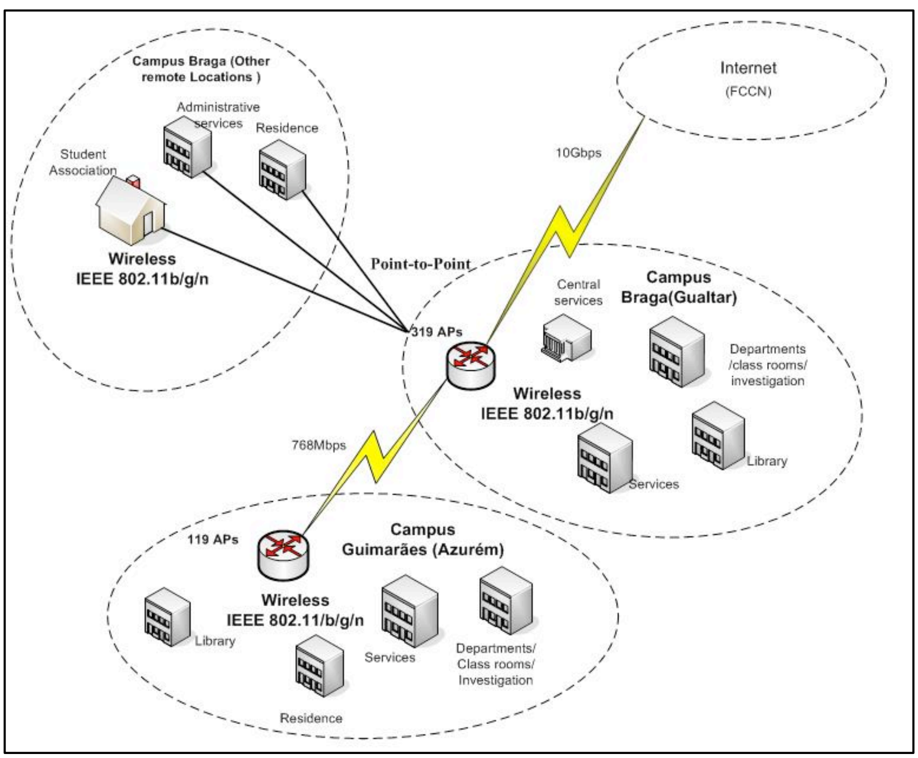

Fig. 1. Partial view of UMinho network infrastructure

A partial view of the UMinho network infrastructure is illustrated in Figure 1. The core of network operation is located in Gualtar, where the main network services are assured to users inside and outside the campus (e.g., in residences, student associations, Azurém campus), providing a 10Gbps access to the Internet (through *Fundação para a Computação Científica Nacional* - FCCN). A 768Mbps link interconnects Gualtar and Azurém campi. The Communications Services headquarters, responsible for the operation and management of the whole network infrastructure and services, are located in Gualtar.

The University of Vigo (UVigo), located in the autonomous community of Galicia, in Spain, was established in 1990. Currently, it has about 30,000 students and approximately 1,800 teachers. The UVigo is located in three cities: Vigo, Ourense and Pontevedra. Each campus has several research centers, schools, residences, sport centers and other service buildings (libraries, administrative offices, canteens, etc.). The WLAN infrastructure is similar to the one at UMinho, diverging on the number of network resources. For instance, Vigo campus has 183 APs distributed across 18 buildings.

4 Data Collection: Challenges and Good Practices

Collecting data in large scale WLANs is in itself a challenge due to aspects such as dense topologies and mobility of users. Taking UMinho eduroam network as an example, the WLAN infrastructure is complex involving hundreds of APs, switches and routers, thousands of users distributed across different buildings and cities, using private or public IP addressing. Therefore, managing and monitoring network operation and collecting traces of WLAN usage is not a straightforward task.

4.1 Data Collection Strategies and Tools

This section provides information on strategies and tools used to assist the data collection process. This process involved the articulation of several technologies, namely:

- *SQL database*: for each AP, the information of authenticated users is stored on an SQL database. This information is directly accessed from a Radius server, which maintains data of eduroam authenticated users. These data include the MAC addresses from wireless devices of authenticated users, and corresponding events, such as association, disassociation, reassociation, roaming, handoff, authentication or deauthentication. Besides, this database allows accounting of multiple records, including the number and duration of user associations, initial and final APs involved in a session, and the traffic volume in packets and bytes for AP. Through an SQL server it is possible to obtain the WLAN traffic history.
- *DHCP logs*: private or public IP addresses are dynamically assigned to wireless devices on campus by a DHCP server. The DHCP logs contain IP addressing assignments to MAC addresses, including a timestamp of that occurrence.
- *SNMP-based tool*: a proprietary tool, developed by the Communications Services, based on PERL and using SNMP primitives, allows monitoring APs at regular intervals (configured for poling each five minutes). This tool performs traffic collection of associated users regardless its authentication status. Network history and statistics can be obtained querying the SQL server. Data can be visualised through a PERL CGI and HTTP-Apache server.
- *Port Mirroring*: for analysis at protocol level, traffic was captured using `tcpdump` via port mirroring. This traffic gathering process was performed at main router providing connectivity between Gualtar and Azurém campi, resorting to Colasoft's Capsa Network Analyzer for traffic analysis.
- *Data Confidentiality*: Processing traffic traces poses confidentiality concerns. It is known that packet header fields (e.g. IP and MAC addresses) or payload allow to access information about network structure, network services location, and user activities. Therefore, privacy policies must be enforced to avoid open processing of network traffic. The use of anonymisation tools (http://www.caida.org/tools/taxonomy/anonymization.xml)

to pre-process network traces, e.g. replacing consistently MAC addresses, allowed to perform WLAN analysis without compromising data confidentiality.

At UVigo, it was used CiscoWorks Wireless LAN Solution Engine (WLSE), which allows managing Cisco aironet APs in WLANs.

Data collection was conducted in two different time periods: the first one took place at UMinho during April and the second one at UVigo, during May/June. The results presented in this paper are mainly focused on eduroam infrastructure at UMinho.

5 Results of Data Analysis

A first step toward WLANs characterisation is to group and classify APs according to their location. University buildings at UMinho are usually associated with a main activity. For instance, the Department of Civil Engineering (DEC) encompasses teacher offices, laboratories and classrooms dedicated to Civil Engineering academic and research activities. APs in this building can be grouped and classified as belonging to DEC. Using this first classification criterion, 30 distinct locations were identified. Additionally, APs were also grouped in type of activity sector, resulting in six distinct sectors, namely Social, Residential, Services, Libraries, Research, and Academic. A third classification criterion was based on the number of distinct users registered in each AP, which leads to the definition of five distinct groups based on the level of APs usage.

The following sections detail the results of characterising and analysing traces from the Azurém Campus, regarding: (i) the number of associations to each AP and corresponding location; (ii) the traffic volumes and the top 10 locations contributing to this parameter; (iii) the average duration of user sessions; (iv) the degree of users mobility; and (v) a fine grain analysis of per day traffic.

5.1 Analysis of associated users and location

Concerning the analysis of WLAN data at UMinho, at first, we have identified the number of different mobile users associated with each of 119 APs (during April). The obtained values ranged from 928 different users on the busiest AP to 2 users on the more underutilised AP. Attending to the asymmetry of WLAN users in using the APs across the campus, and to provide a finer analysis of the utilisation patterns, five distinct activity groups were considered. As illustrated below, Group 1 comprises the total of APs with 500 or more users, illustrating zones of higher user preference:

```
Group 1: APs with number of users ≥ 500
Group 2: APs with 300 ≤ number of users < 500
Group 3: APs with 100 ≤ number of users < 300
Group 4: APs with 50 ≤ number of users < 100
Group 5: APs with 0 ≤ number of users < 50
```

Considering these intervals, we found that AP utilisation follows approximately a normal distribution centered on Group 3 with 38% of AP associations. A total of 13.4% APs handle more than 500 users, whereas around 16% handle less that 50 users. This analysis also indicates that 56% of APs on campus support between 100 and 500 users. Figure 2a) shows the values recorded for the different intervals.

Taking into account the results above, for each group, the spatial distribution of APs on campus was assessed. Figure 2b) illustrates the obtained results when considering 30 spatial areas (sites). As shown, the busiest APs cover about 30% of the total area, while less utilised APs (Groups 4 and 5) around 30%.

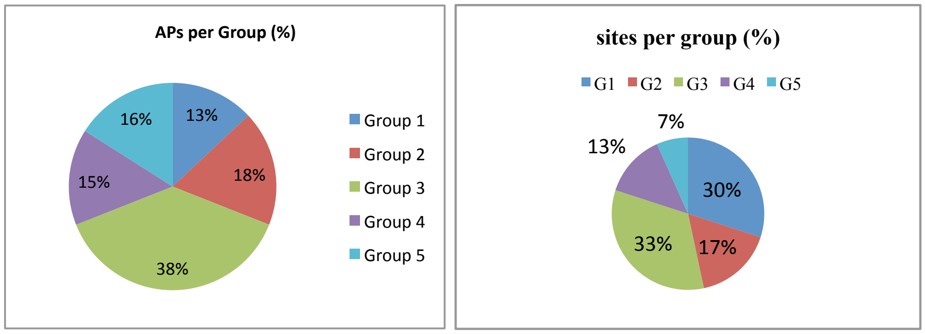

Fig. 2. a) Distribution of APs per group; b) Distribution of sites per group (%)

5.2 Analysis of traffic volume

The analysis of WLAN traffic volume (eduroam) on campus was carried out considering the APs grouped in six classes, according to their location:

(i) Social: in bars, canteens and sports facilities;

(ii) Residential: in student accommodation premises;

(iii) Services: in administrative, technical and student support services;

(iv) Library: in libraries;

(v) Research: in research laboratories;

(vi) Academic: in departments and schools.

Figure 3 shows the distribution of the volume of inbound and outbound traffic by type of location, with a clear dominance of inbound traffic. As illustrated, the residential and academic sectors are responsible for most of inbound traffic, with nearly 80% of the total. Regarding the volume of outbound traffic, residential and academic sectors are again the major contributors to the overall traffic load. In absolute terms, for the period under study, the traffic generated by eduroam users was 620GB and 5120GB of outbound and inbound traffic, respectively. In order to detail the analysis of traffic volumes, accounting data was processed to determine the location of most active APs.

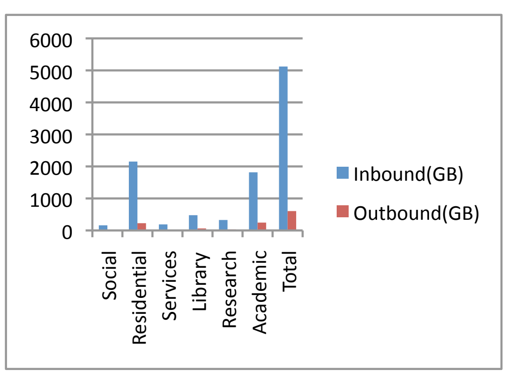

Fig. 3. Distribution of traffic volume by activity sector

5.3 Traffic Volume: Top 10 APs

The Top 10 ranking identifies the 10 APs exhibiting higher traffic volume. These APs represent approximately 2.3 TB of inbound traffic, i.e., 45% of the total traffic. As shown in Figure 4, in the Top 10 group, 50% of APs belong to Residential sector and 30% to Library sector. The Library APs are also part of group 1, i.e., group with the largest number of associated users. This Top 10 also includes APs belonging to Services and Research.

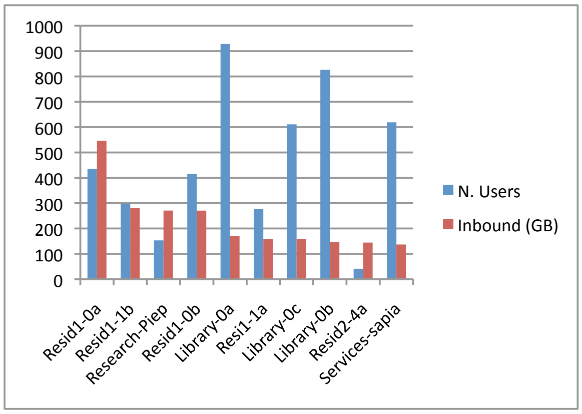

Fig. 4. Top10 - Traffic volume per AP

5.4 Analysis of sessions

An important aspect of users' pattern behaviour is the time users remain associated to each AP. This time analysis, based on the duration of user sessions, resorts to aggregated data comprising the total number of sessions and their average duration. Crossing this information with the number of users registered per AP, it is possible to evaluate the average time of each user session. Based on this calculation, it was observed that the AP with the highest monthly associated time per user was approximately 58 hours, resulting from the sum of its individual sessions.

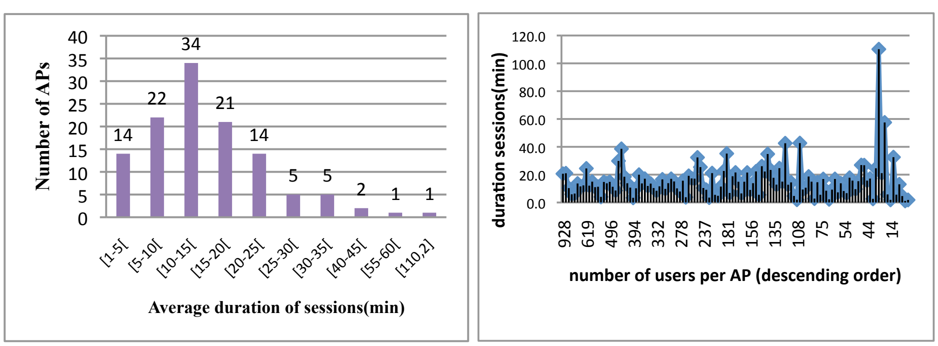

Fig. 5. a) Number of APs per average time duration of sessions; b) Distribution of duration of sessions

The average duration of sessions in APs varies between 1 and 110 minutes. Figure 5a) shows a histogram representing the number of APs for distinct session time intervals. As illustrated, five representative time intervals were identified, showing that, from a total of 119 APs, 34 (28%) handle sessions with average duration from 10 to 15 minutes. Based on this information it was found that: (i) APs with a low number of users (Groups 4 and 5) support a large number and duration of sessions. These observations correspond mainly to APs associated with the Residential sector and other APs in locations of reduced mobility; (ii) the average length of sessions does not exhibit major variations, especially in groups 1 and 2. The largest deviation occurs in all AP with few users. There are sessions of high duration depending on the AP location (e.g. home or research center) and other APs with many users and sessions of reduced duration. Figure 5b) shows the distribution of the average duration of sessions per AP in descending order of number of users.

5.5 Daily Analysis

A fine grain analysis of WLAN traffic is provided through the identification of users' behaviour and patterns in small timescales. Taking a weekly traffic trace (from April 2010) as an example, traffic data was analysed on a daily basis, hour-by-hour. This section reports the results obtained regarding: (i) the variation on number of users and on number of sessions; (ii) the aggregated traffic volume and packet size characteristics, for time periods of low, medium and high traffic load.

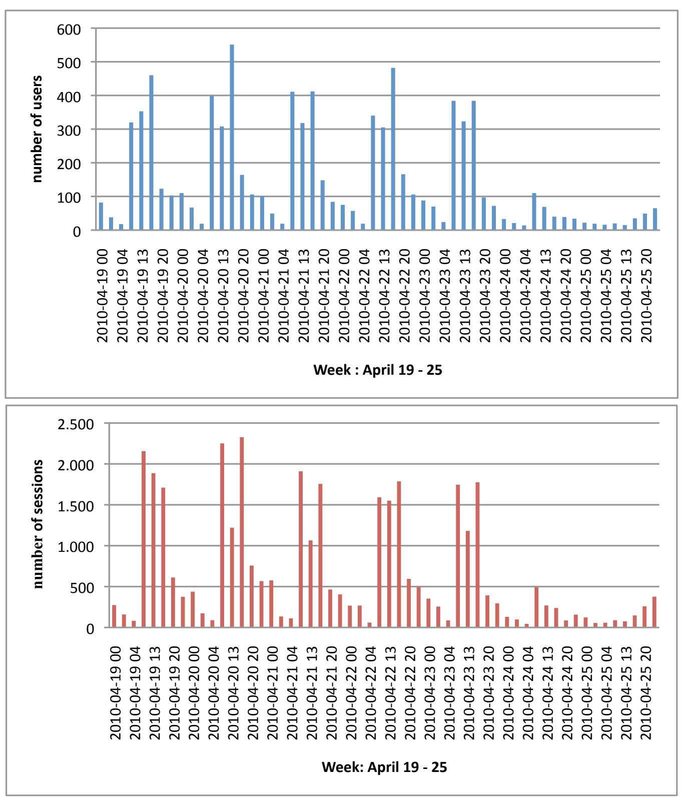

Fig. 6. a) Variation on the number of users; b) Variation on the number of sessions

Users and Sessions Accounting As illustrated in Figure 6a), the variation of the number of users during working days follows a pattern with a daily peak at 11/12 and 16 pm, with about 400-550 wireless users. The low activity period occurs at night, with a minimum number of users at 4 am (about 15-20 users), mainly from student residences. The variation on the number of sessions presents a similar behaviour. As illustrated in Figure 6b), during busy hours, the average number of sessions per user is around 4 (about 2350 sessions for 550 users), increasing to 7 overnight (e.g., about 100 sessions for 15 users).

Traffic Volume and Packet Length The analysis of inbound traffic volume between Gualtar-Azurém during working days exhibits two peak hours and three activity time periods. As shown, the peak hours occur at noon and 4pm; the time period with high traffic load occurs between 9am to 6pm, reaching an inbound rate of 100Mbps; from 6pm to midnight this rate decreases to 40Mbps, keeping this level up to 9am. Regarding the link capacity between the two campi, WLAN traffic volumes represent 4% to 10% of the available capacity.

Packet size can be an important aspect on the performance of network equipment, with special impact on queuing behaviour. The analysis at packet level within the WLAN reveals two main groups of packet length: centred on 64 Bytes and 1518 Bytes, as illustrated in Figure 7. While smaller packets represent 35% of the total number of packets and 2,76% of the total number of bytes transmitted, larger packets represents 40% of the total packet count and more than 75% of the bytes transmitted.

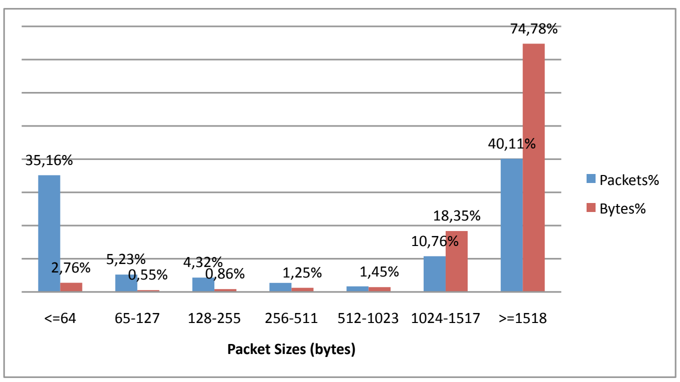

Fig. 7. Packet sizes and bytes (%)

5.6 Mobility of Users

Allowing user mobility is considered one of the main advantages of WLANs. On campus, users move to different locations, e.g. classrooms, canteens, accommodations, sport centers, meaning that mobile devices should necessarily go through the process of association with at least two APs.

To assess to what extent users take advantage of mobility in the WLAN, we identified (by MAC address) the users that presented high mobility patterns (during April and May). In this way, we analysed the number of associations from each user with different APs. The results show that most of authenticated users (90% of a total of 3480 users) exhibited effective mobility and only a reduced percentage (9%) used a single AP to access the WLAN.

Attending to the high percentage of users associated with two or more distinct APs, a more detailed analysis was carried out, identifying the number of users per mobility scenario, grouped in intervals of number of associated APs. As shown in the histogram of Figure 8, a significant number of users (around 37,5% of the total) visited from 2 to 5 APs. The tail of the histogram indicates that users may reach a maximum of [58,61] distinct APs. This means that, monthly, a large number of different locations on campus may be visited per a single user. Figure 9 provides a complementary view of the mobility degree expressing the relative values obtained in April. For users with the highest mobility indicators (corresponding to the slice "other ranges"), approximately 37% were identified as the same users in April and May.

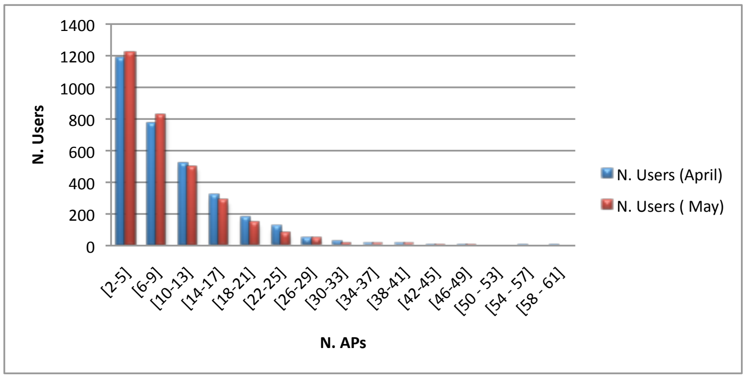

Fig. 8. Users mobility

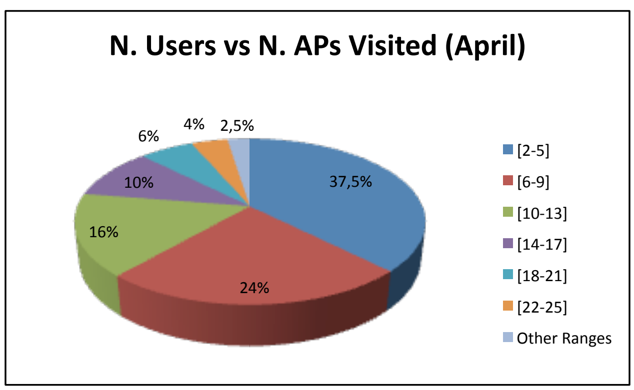

Fig. 9. Relative view of users mobility

5.7 UVigo WLAN Considerations

This section presents briefly the results from analysing WLAN traffic at UVigo. The study was focused on Vigo Campus, which encompasses 183 APs distributed across 18 buildings. As in UMinho, in order to identify the more relevant sites for analysis, the number of registered associations in the 183 APs during May/June was analysed. From the more representative sites regarding WLAN usage, it is clear that libraries, social areas and free spaces on campus are the preferred locations for students to access the wireless network. Although the quantification of results is clearly distinct due to the number of APs and users involved, the patterns of WLAN usage follow closely the ones reported in this paper for UMinho WLANs.

6 Conclusions

This paper has focused on the characterisation of University WLANs usage, within the eduroam initiative. The study, based on real traffic traces collected from UMinho and UVigo WLANs, contributes to understand to what extent users take advantage of this technology, within the academic context. The study has covered and assessed several network usage metrics related to AP utilisation, session characterisation, access location influence, and user mobility patterns. These metrics are relevant to better understand both the real utilisation of the WLAN infrastructure and the characteristics in terms of user behaviour.

Results from traffic analysis showed that the number of users associated with an AP depends heavily on the activity sector where those APs are located (e.g. libraries). The number and duration of sessions also varies depending on location and type of activity. In fact, the duration of sessions is more related to the activity area where APs are located than with the number of associated users. As regards the overall traffic volume, both residential and academic sectors are clearly the main contributors. The mobility analysis reveals that a significant number of users access the WLAN from a reduced number of APs in campus, exhibiting a stable behaviour in location preference. However, the mobility indicators also reveal that a considerable number of users use a large number of different locations on campus for accessing the network.

The current results, based on recent data and focused on user behaviour, constitute a useful feedback for planning, dimensioning and managing WLAN within eduroam context. Future work includes expending the characterisation of traffic on UMinho WLAN, including aspects such as the study of mobility patterns, protocols and applications per activity sector and user location.

Acknowledgements The authors would like to thank the Communications Services of the University of Minho and University of Vigo for the technical support provided.

References

1. eduroam: Eduroam Initiative, http://www.eduroam.org (2003)
2. Tang, D., Baker, M.: Analysis of a local-area wireless network. In: MOBICOM. (2000) 1–10
3. Kotz, D., Essien, K.: Analysis of a Campus-wide Wireless Network. In: In Proceedings of ACM Mobicom, ACM Press (2002) 107–118
4. Henderson, T., Kotz, D., Abyzov, I.: The Changing Usage of a Mature Campus-Wide Wireless Network. In: MOBICOM'04. (2004)
5. Balazinska, M., Castro, P.: Characterizing mobility and network usage in a corporate wireless local-area network. In: Proceedings of the 1st international conference on Mobile systems, applications and services. MobiSys '03, New York, NY, USA, ACM (2003) 303–316
6. Schwab, D., Bunt, R.: Characterising the Use of a Campus Wireless Network. In: INFOCOM. (2004)
7. Papadopouli, M., Moudatsos, M., Karaliopoulos, M.: Modeling Roaming in Large-scaleWireless Networks Using Real Measurements. In: Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks. WOWMOM '06, Washington, DC, USA, IEEE Computer Society (2006) 536–541
8. Kumar, U., Yadav, N., Helmy, A.: Gender-based feature analysis in campus-wide WLANs. SIGMOBILE Mobile Computer Communications Review **12** (2008) 40–42
9. Kim, J., Helmy, A.: Analyzing mobility evolution in WLAN users: how predictable are we? SIGMOBILE Mobile Computer Communications Review **14** (2010) 10–12
10. Gast, M.: 802.11 Wireless Networks: The Definitive Guide. O'Reilly, 2nd Edition (2005)
11. IEEE: IEEE Std. 802.11-1999, Part 11: Wireless LAN Medium AccessControl (MAC) and Physical Layer (PHY) specifications, Reference number ISO/IEC 8802-11:1999(E) (1999)