Revisiting the Modeling of User Association Patterns in a University Wireless Network

Ljubica Pajevic[‖*], Viktoria Fodor[*] and Gunnar Karlsson[*]

‖Department of Informatics, Technical University of Munich, Munich, Germany

*Department of Network and Systems Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

{ljubica, vjfodor, gk}@kth.se

***Abstract*—This paper presents an analysis of a large trace of user associations in a university wireless network, which includes around one thousand access points over five campuses. The trace is obtained from RADIUS authentication logs and its merit is in its recency, scale and duration. We propose a methodology for extracting association statistics from these logs, and look at visiting time distributions and processes of user arrivals to access points. We find that a large fraction of the network—around half of all access points—experiences time-varying Poisson arrival process, and association distributions can be modeled by two-stage hyper-exponential distributions at most of the access point. While network associations in campus wireless networks have been extensively studied in the literature, our study reveals changing patterns in user arrival processes and association durations, which seem to be characteristic for networks of predominantly mobile users, and allows the use of tractable network occupancy models.**

***Index Terms*—WLAN, user mobility, trace-collection analysis.**

I. Introduction

The ubiquity of wireless local area networks (WLANs) in today's workplaces, universities, and other public areas, increased the importance of the accurate characterization of users' access and usage patterns. Such characterization is essential for network dimensioning and capacity planning, and can support the design of efficient network algorithms, e.g. for energy saving [1], or content caching [2]. While similar studies have been done before [3], [4], [5], [6], [7], the proliferation of mobile devices and the densification of the networks make new evaluations necessary.

In this study we re-evaluate whether user associations can be modeled by tractable Markovian models under these new circumstances, utilizing traces of network associations in a university campus wireless network based on Eduroam association records. Our main contributions are the following. First, we propose a methodology to use the RADIUS logs, that are easy to acquire and are often available in production networks, for the analysis of user association patterns. Second, as we analyze a recent and relatively large trace, we can identify the effects of the current trends in network usage. We provide a thorough evaluation of the Poisson arrival hypothesis, and identify the cases when it needs to be rejected. Furthermore, we show that connection times can be fitted to relatively simple, two-stage hyper-exponential distributions, thanks to the high number of connected mobile devices.

In the remainder of the paper, we first give an overview of the wireless network in question and the data collection process, in Section II. Section III details the trace processing methodology. The results of our analysis and modeling are presented in Section IV. We discuss related work in Section V, and conclude our study in Section VI.

II. Data collection

A. Acquisition of the authentication events dataset

The data trace comes from the campus wireless network of KTH Royal Institute of Technology. The KTH WLAN provides coverage for buildings on one large and four small campuses located within metropolitan Stockholm area. The campus buildings are for non-residential use, housing classrooms, computer laboratories, libraries, offices, administrative premises, cafeterias and restaurants. The largest site, the Main Campus, includes 48 buildings with around 790 access points (APs). Due to the high density of APs and proximity of campus buildings, most of the outdoor areas are covered as well. All APs in the network are Cisco Aironet models. At the time of trace collection (2014–2015) the university had around 18000 active students and employees, most of them accessing the wireless network via smartphones, laptops and other portable devices. The number of active wireless users that connected to the network for a weekday varied between 13000 and 15000.

The raw dataset consists of Eduroam associations collected from the authentication server deployed in the university network. These associations constitute 95 % of all associations in the wireless network. User association in an Eduroam network requires RADIUS authentication. The basic network architecture is depicted in the diagram in Fig. 1.

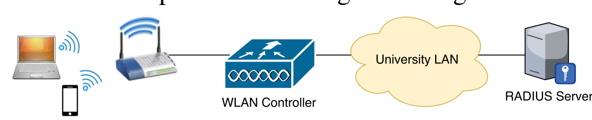


Fig. 1. WLAN network diagram with authentication server.

The APs are managed by Cisco 5508 Series Wireless Controllers (WLC). The central component in the secured WLAN environment is the authentication server; it handles user access requests either by processing requests itself or by proxying them to another server in the user's home institution. The authentication server in the network is FreeRADIUS server. For keeping record of all authentication events, FreeRADIUS uses a syslog module *linelog* which reports events in the F-Ticks format [8]. Each line corresponds to an authentication event, called a *tick*, containing information about the user device requesting authentication, the timestamp of the event and the result—access accepted or rejected—as well as the name (identifier) and MAC address of the access point. These records are logged and reported to the national operator for collecting statistics at national levels. Prior to sending the records, client MAC addresses are anonymized through hashing, which obfuscates the user's identity, though keeps the hashed address preserved throughout the trace. The trace does not contain the exact locations of the users (e.g., GPS coordinates) but it gives the identity of the access point that the client associated with.

Our goal is to extract information about user connectivity patterns over the observed time period. It is clear that determining such patterns from association events only—without the complementary information of de-authentication—leads to uncertainty. Therefore we resort to reconstructing user's associations to the network and the corresponding connection time by utilizing the knowledge of the association process and empirical measurements from the client side. In addition, we performed *war-walking* to estimate the outdoor wireless transmission range, and to discover areas not covered by any AP and areas covered by several APs, which are fundamental to determine the connection times of the mobile users moving out from an AP area. The results of our measurements in the Main Campus area are shown on Fig. 2. The size of the points represents the signal strength, the opacity of the circles represents the accuracy of the estimated GPS location, while different colors correspond to different buildings on the map.

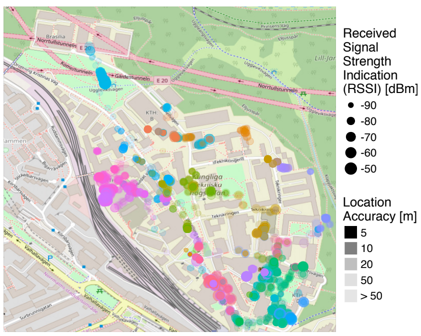


Fig. 2. Coverage and received signal strength indication (RSSI) of access points mapped by *war-walking*.

B. Details of RADIUS authentication

The access points in the network support IEEE 802.1X authentication standard. 802.1X encompasses the use of the Extensible Authentication Protocol (EAP), a container protocol that carries the actual authentication data inside. Two authentication scenarios can be distinguished. In the first scenario, a user device joins the network by associating with an AP. An EAP tunnel is established between the user and the server, enabling the two end entities to mutually authenticate each other. Then, the authenticator (RADIUS server) issues a challenge for the user, and grants access or rejects the request. Entries in our dataset correspond to *access granted* events. The server next closes the EAP tunnel and sends the key-generating material to AP. Only the AP and the user participate in the remaining authentication steps. The other case of authentication takes place when a client is already connected to an AP in the network and roams to a different AP. In a wireless network which does not support fast-secure roaming protocols—this is our case—the client must go through the exact same process and perform a full authentication. In addition, the following times trigger the re-authentication of the users:

- *User idle timeout,* T_{idle}: When a user is idle without any communication with the AP for the amount of time set by this value, the client gets de-authenticated by the WLC, and it must re-authenticate and re-associate to the WLC. The default value of the user idle timeout is 5 minutes, which is confirmed by the client side measurements.
- *Session timeout*, T_{sess}: As a part of a security precaution, all active users need to re-authenticate at regular time intervals, defined by the session-timeout value. The default value is 30 minutes, which is also confirmed by the RADIUS log.
- *Periodic scanning,* T_{scan}: Mobile devices perform periodic scanning every 10–15 minutes, to find the AP with the best wireless connection, and in every 5 minutes if they are in idle state. The periodic scanning leads to re-associations to the same AP, if the wireless channel conditions did not change. Here we consider T_{scan} = 5 minutes.
- *Broadcast key interval*: When a user connects to the WLAN, he receives a broadcast key, that allows encryption of broadcast/multicast traffic. To protect from attackers, the AP pushes group keys to the clients at scheduled times, the default interval is one hour.
- *Device timeout*: All wireless user devices have their own settings for determining when and how to disassociate from a WLAN, or disable the wireless network card. The default values are device-dependent.

III. Initial trace processing

The dataset we collected spans over 16 months, during which the number of active APs deployed in the network varied from 934 to 985. For the purpose of this study, we choose a subset of four term weeks, from September 22 to October 19, 2014, when the network topology was stable and the user behaviour was typical for the university network. The total number of ticks during this period was 7040095, comprising around 300–350 thousand lines for week days and 60–80 thousand lines for weekends, capturing 35649 unique devices. Around 42% of the devices were active ten or more days, and 8% of the devices appeared in the trace with just a single association event.

As the available trace contains only *(re-)association* events, it does not give direct information on how long a user has been connected to an AP. Therefore, we derive heuristics to estimate the connection times. The heuristics utilize the authentication state machine, the wireless coverage map, as well as the knowledge of the key timer values from Section II-B. In addition, we consider the following time intervals:

- T_{auth}, the time required for the RADIUS authentication process. Our client side measurements suggest $T_{auth} = 1$second.
- T_{min}, the minimum association time when a device has started moving away from the AP. Based on the coverage map, we consider $T_{min} = 30$ seconds.
- $T_{roam} = T_{idle} + T_{scan} = 10$ minutes, the maximum interval between consecutive associations under roaming, that is, when the user is on the move. The maximum time happens when the user is idle.
- t_d denotes the time difference of the consecutive association events for a user, that is, the *association interval*, as recorded in the RADIUS trace.

For each user in the trace, we perform parsing of the raw trace on a time window of one day to extract the user's trajectory defined by the (*AP, visiting time*) tuples. The parsing includes the merging of successive associations, the removal of short associations, and the estimation of the connection times. Here are our rationales.

1) Merging successive associations with the same AP: Users associate and re-associate with the same AP after user idle timeout, after session timeout, or for mobile devices due to periodic scanning. Observing a sequence of associations with the same AP, within one hour, we assume that the client has not moved away from the AP and we merge all the consecutive associations into a single association.

2) Removal of short associations: Occasionally, multiple association events with the same or nearly the same timestamp are logged. Since this is not enough time to complete the post-authentication steps, we ignore association intervals $t_d \leq T_{auth}$ and keep only the last entry in the series of short associations.

3) Estimation of unknown connection times: Consecutive association events that are more than an hour apart or involve APs that are not neighbors[1] on the coverage map suggest that the user was disconnected between the two associations, and we estimate the connection time before disconnection as follows.

i) Short disconnections occur when consecutive associations happen at non-neighboring APs, and the association interval is longer than T_{min} and shorter than T_{roam}. In this case we select the association time randomly uniformly in the interval $[T_{min}, \min(t_d, T_{idle})]$.

ii) Medium disconnections happen when $T_{roam} < t_d \leq T_{sess}$, with successive associations at non-neighboring APs. Since a new association eventually happens within a session timeout time, the user is likely to be moving in the campus area, experiencing longer disconnection time. We calculate the association time based on the wireless coverage map, by subtracting the estimated walking time from the association interval. For an estimation of the walking time we use Google Distance Matrix API [9].

iii) Long disconnections occur when $t_d > T_{sess}$. This is the case when the user is disconnected for a longer time, or even leaves the area. We estimate the length of the last association time based on the previous association interval. If the previous association was longer than T_{roam}, the user is likely to be static, and we select the association time uniformly in random from the interval of $[T_{roam}, T_{sess}]$. Otherwise the user is likely to be on the move, and we assign some random association time from $[T_{idle}, T_{roam}]$.

Out of 7040095 association events in the four-week trace, around 4% of the entries where short and another 4% long disconnections, whereas the medium disconnections were observed less frequently, amounting to 1% in total.

IV. Access Pattern Modeling and Validation

In this section, we present the main findings of our study. We characterize access patterns of users at each individual AP, first with respect to the arrival processes and subsequently by modeling the user visiting times. Our main objective is to evaluate the possibility to model association patterns in a Markovian framework, that is, considering Poisson arrival process and phase-type visiting time distribution.

Since the network exhibits very low activity during weekends we analyze only work days in the set—a subset of 20 days of the initial 28 days—and consider the arrivals between 8 AM and 6 PM, excluding the daily sets when the number of data points (arrivals) is smaller than 20. We also exclude arrivals that occur during the night and stay during the day, and the opposite—arrivals that occur during the day and do not leave during the night, as they are re-association events generated by stationary users. The fraction of the discarded, stationary users ranges from 2% to 6% per day, or 3.6% on average, which equals 521 users. The resulting set contains 913 access points having at least one day of arrivals subjected to modeling.

A. Arrival process modeling and validation

Clearly, the intensity of arrivals to an AP changes in time, and therefore, as earlier studies [4], [5], [6], herein we evaluate whether arrivals can be modeled with a non-homogeneous (also denoted as *time-varying*) Poisson process. The counting process $\{N(t); t \geq 0\}$ is said to be a non-stationary or non-homogeneous Poisson process with time-varying arrival rate $\lambda(t), t \geq 0$ if: *i)* $N(0){=}0$, *ii)* $N(t)$ has independent increments, and *iii)* $N(t){-}N(s) \sim Poisson\left(\int_s^t \lambda(u)du\right)$ for $s < t$.

This arrival process retains the property of independent and exponentially distributed inter-arrival times. It should be noted that the studies related to ours rarely consider both of these requirements, usually focusing only on the assumption of exponential distribution, [5], [6]. Below we provide a procedure that considers both the distribution and the independence of the inter-arrival time samples.

The estimation of the rate function is often performed by dividing the sample sequence—either inter-arrival times or arrival count—into short time intervals and assuming that the rate is constant during those intervals [4], [6]. We approach the modeling task from a different angle, assuming that the rate function of the Poisson process varies more slowly—in order of hours. To this end, we deploy a *change point detection* (CPD) algorithm [10] to split the sample sequence of inter-arrival times into segments for which we postulate that the process behaves as a homogeneous Poisson one. The CPD algorithm searches for abrupt changes in the mean and/or the variance of the sample sequence. The main parameter of the algorithm is a threshold value, that ensures a constant probability of a false positive occurring after each observation, determining the *Average run length* (ARL) as the expected number of observations received before a change is falsely detected. For sequence processing we use the implementation of Ross [10], available in the *Change point model* (`cpm`) package [11], and we consider two ARL values, ARL=500 and 5000, to see the effect of how conservative the algorithm is.

Fig. 3 gives an example on how the CPD algorithm divides a single day inter-arrival times observed at one AP (located in the main library) into disjoint segments with different intensities: in the top figure, the (yellow) dots represent the inter-arrival times with the $x-$axis marking the time when the

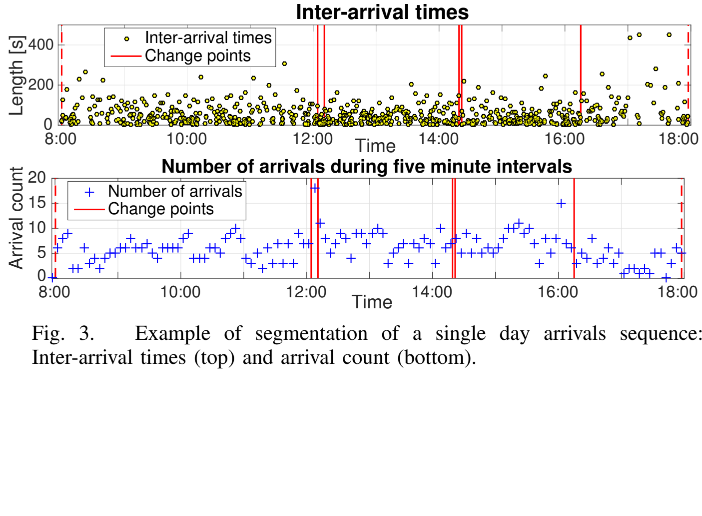
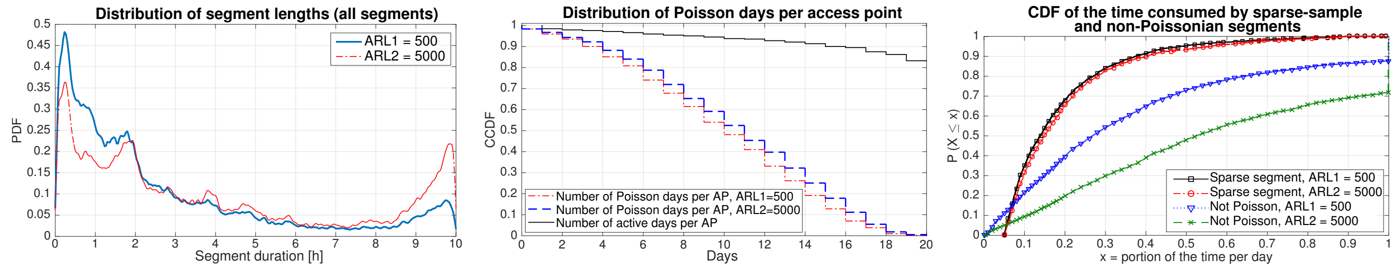


Fig. 3. Example of segmentation of a single day arrivals sequence: Inter-arrival times (top) and arrival count (bottom).

[1] We consider that two APs are neighboring if they are located in the same building or, otherwise, in two adjacent buildings not more than 100 m apart.



(a) (b) (c)

Fig. 4. (a) Probability density function of the segment duration. (b) CCDF of the number of days per AP passing the test for non-homogeneous Poisson arrivals. (c) CDF of the portion of time per day consumed by sparse sample segments and non-Poisonian segments.

TABLE I. STATISTICS FOR ENTIRE DAYS

Days count; fraction of all days	ARL $ARL1 = 500$	$ARL2 = 5000$
Pass test	9068 (49.66 %)	9685 (53.04 %)
Long sparse sample	1915 (10.49 %)	1033 (5.66 %)
Not Poissonian	6078 (33.28 %)	6343 (34.74 %)
Sparse sample—entire day	1199 (6.56 %)	
Total # days	18260	

TABLE II. STATISTICS FOR SEGMENTS

Segment categories	**Total # segments**		**Average duration [minutes]**	
	$ARL1$	$ARL2$	$ARL1$	$ARL2$
Poisson	50990	34348	158.9	217.4
Independent and not Exp.	4048	4845	184.7	284.1
Exp., and dependent	3624	2716	135.2	209.1
Not Exp. NOR independent	283	400	137.8	281.0
Sparse sample segments	12962	5167	27.5	40.3

sample was generated; the vertical (red) lines correspond to the time instances of the detected change point. A better insight into this segmentation can be obtained from the lower plot in Fig. 3, which depicts arrival counts in five minute intervals: the vertical lines around 12 AM clearly mark a burst of arrivals.

While the applied algorithm assumes that samples are independent and exponentially distributed, we need to verify that this holds, hence we investigate each of the segments independently using two tests. First, for each set of inter-arrival times we perform Kolmogorov-Smirnov (KS) testing with modified statistics table, that is, the Lilliefors test for exponentiality [12] (corresponding to the 5% significance level), to test whether the samples come from an exponential distribution with the rate parameter estimated from the set.

For testing the independence of the inter-arrival sequences in the generated segments, we use Brock-Dechert-Scheinkman (BDS) test [13], which tests the null hypothesis that data comes from a process that generates independent and identically distributed (i.i.d) samples. The BDS test embeds the observed sequence in high dimensional vectors, and the dependence is examined by counting "near" points in space. We follow the modified testing procedure for small sample sizes (which is our case) given in [14]. Specifically, we consider dimensions $1, \ldots, 5$, a distance limit of one standard deviation of the tested sample, and a significance level of 5%.

B. Numerical results on arrival process modeling

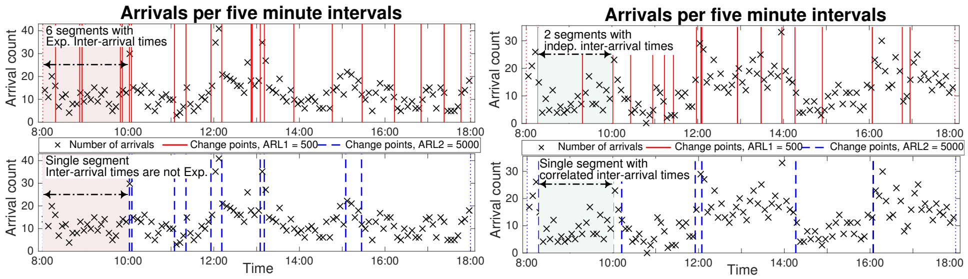
In this section we evaluate the hypothesis that arrivals can be modeled with a time-varying Poisson arrival process. For change point detection we consider two values, ARL = 500 and 5000. We evaluate the segment durations to see whether they are long enough for further analysis. The two segmentation procedures produce, respectively, 71888 and 47482 segments in total—over all days and all APs—with, on average, 4.21 and 2.78 segments per day. The distribution of segment lengths is shown in Fig. 4(a). We see several characteristic peaks, due to authentication timers, as well as the typical schedule of lectures at the university. The peak at 10 hours represent APs with stationary arrival process during the observed working hours.

1) Assessment of the Poisson arrival process hypothesis: First we evaluate how often the assumptions of exponential and independent inter-arrival times are met. We test only the segments with more than ten samples, labeled as *regular*, since segments comprising ten or fewer samples, denoted as *sparse sample segments*, are too small for sound statistical analysis. We consider that the entire day passes the test for time-varying Poisson arrivals if all regular segments pass the test, and sparse sample segments amount up to less that 5% of the AP's daily active time.

If we consider all APs on all days, only around half of the entire-day sequences pass the test, specifically 50 or 53%, for the two ARL values. Fig. 4(b) shows the complementary cumulative distribution function (CCDF) of the number of days the APs pass the Poisson test. The figure shows as well the distribution of active APs, and reflects that the number of days discarded due to less than 20 samples is not negligible: only 760 APs, or 86% are considered active all 20 days. We observe that the ARL value does not significantly affect the results. The large gap between the two lower CCDF stair-plots and the tail distribution of active APs clearly indicates that many of the APs fail the test of the Poisson arrivals a large number of days. For example, only 413 APs (less than half) observe Poisson arrivals for more than ten days, and 148 (less than 20%) more than 15 days. Some APs, in fact, rarely perform in accordance with the assumed arrival process: 23 APs never pass the test, and 242 APs pass only for 5 days or less.

2) Examining the structure of intervals that are not Poissonian: Table I gives a quantitative comparison of the reason of failing the test: 1) the total time duration of small-size segments exceeds 5%—we label the days falling into this category as days with *long sparse sample* (*LSS*) segments or 2) at least one of the segments in a day does not exhibit homogeneous Poisson arrival process—the day contains *not-Poissonian* (NP) segments. We see that the main reason of rejection (for ca. 34 % of the days) is that latter one, that is, the day contains NP segments.

To see how dominant sparse sample or non-Poissonian segments are, Fig. 4(c) shows the CDF of the portion of time a day these non-conforming segments cover, considering



(a) Independent and not exponential samples.

(b) Exponential and dependent samples.

Fig. 5. Examples of intervals failing the exponential distribution or the independence test.

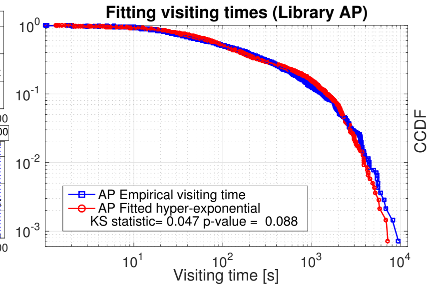


Fig. 6. Visiting time distribution fitted with H_2.

the days when these segments are present. We see that the ARL value does not significantly affect the detected small size samples interval, and the LSS segments in general cover a small part of the day only. On average, considering all LSS days, LSS segments contribute to less than 20% of the day. Looking at the time span of non-Poissonian segment, we see that it is significantly affected by the ARL value: higher ARL means less frequent change detection, and as a result, longer non-Poissonian segments. Non-Poisson segments can be rather dominant in a day, the average portion of the time consumed by NP segments is 37 respective 57%.

By looking at NP segments only, next we analyze which of the conditions—non-exponential or dependent inter-arrival times—is more likely to reject the hypothesis. As Table II shows, non-exponential distribution is the main reason for rejection, but overall, the two categories are comparable with respect to the number, and the average lengths of the segments.

Finally, we demonstrate the effect of the ARL parameter in Fig. 5. We select a busy AP in one of the main buildings of the Main Campus, that has also large outdoor coverage. The usual number of users that associate with this AP is between 1100–1200 per day. Fig. 5(a) shows the arrival count per five minute intervals and the time instances of change points for two ARL values. From 8 AM to around 10 AM there are 269 samples and the longer segment, under ARL2 = 5000 (lower figure) fails the test for exponential inter-arrivals. As opposed to this, when ARL1 = 500, the same time interval is split into six segments (figure above) and apart from the one small segment, comprising a batch arrival of 10 users within 161 seconds, all other segments pass the exponential distribution test. A similar example, from another day, is illustrated in Fig. 5(b). The second segment in the lower plot, from 10 AM to 12 AM, fails the test of independence, showing a weak linear dependence among inter-arrival times. The lower ARL value (figure above) splits the interval into half, and breaks the dependence. As a conclusion, we see that the ARL value needs to be selected with some care, to balance between rejecting the hypothesis due to too long sequences, and generating segments that contain two few samples.

C. Analysis of the visiting time distribution

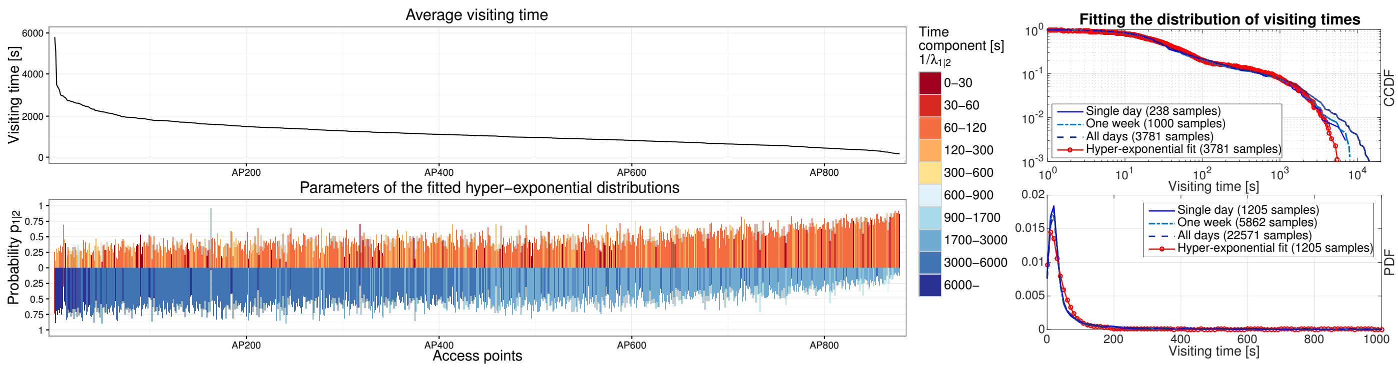
In this section we study the visiting time duration which we define as a continuous time that the user is connected to the same AP. Again, we consider only sessions that are limited in time by the work day window of 10 hours, and we characterize the visiting times at separate APs.

Previously established results indicate that the visiting time at APs in a campus network is long-tailed [3], [5], [7]. As shown in [15], phase-type distributions, and specifically, the hyper-exponential distribution, provide suitable approximation. After testing different number of phases, we found that for our case the two-phase, that is, the H_2 distribution provides the best goodness-to-fit statistics. That is, we fit the visiting time duration at an AP with the function $f_{H_2} = p_1 f_{Y_1} + p_2 f_{Y_2}$, where $p_1 + p_2 = 1$ and $Y_{1|2}$ are exponentially distributed random variables with parameter $\lambda_{1|2}$.

We apply the expectation-maximization (EM) algorithm to estimate the parameters $p_{1|2}, \lambda_{1|2}$. To ensure a sufficient sample size, still allowing the EM algorithm to converge, we use visiting time samples from a single week. For APs with very sparse samples we consider all four weeks, while for some very busy ones only a single day. We do not model 17 APs with less than 50 arrivals in the four weeks.

We find that at 880 out of 896 APs the visiting time distribution can be modeled with H_2 distribution with high accuracy, as exemplified in Fig. 6. The average visiting times for these APs are plotted in the top graph of Fig. 7(a), with the parameters of the H_2 distributions shown in the bottom graph: the $p_{1|2}$ values are represented by the heights of the bars and $1/\lambda_{1|2}$ values are color coded. We see that the connections can be split in two groups, one with very short visiting times, in the range of minutes, and another one with long visiting times, up to a couple of hours. Short visiting times clearly correspond to associations when the user is likely on-the-move or the device is searching for a stable connection. Short associations are present at all APs, often with high probability, which shows the large penetration of handhold mobile devices.

The third group, counting 16 APs, observes visiting times that can not be fitted with the proposed H_2. Looking at the plots of these distributions, we can identify three possible reasons. First, a mismatch can be caused by excessively long visiting times, which the exponential components do not capture. An example is given in Fig. 7(b) (top) which compares fitting of the empirical visiting time distribution, representing samples of a single day, one-week and all-days associations. The CCDFs of empirical and fitted H_2 show that the fit for this AP has a shorter tail than the empirical distribution. The second reason for poor fitting is the occurrence of very short associations; these associations are shorter than 20 seconds and result from the mobility of users in a dense AP deployment. Fig. 7(b) (bottom) shows the probability density of visiting time distribution that falls into this category. Finally, a few APs with many, dominantly idle users fail due to smaller peaks at multiples of T_{idle}, while showing good fit otherwise.



(a) (b)

Fig. 7. Visiting time distribution. (a) Parameters of the fitted hyper-exponential distributions. (b) Examples of the visiting time not conforming to the H_2 fit.

V. Comparison with previous results

User access and association patterns have been investigated both in university [3], [6], [7], [4], [16] and in urban WLAN scenarios [5], [6], with network traces from 2003 to 2011, demonstrating the changes in the network usage patterns. The time-varying Poisson arrival process has been shown to be accurate in many scenarios in [4], [5], [6]. In this work we evaluate the exceptions in detail, and conclude that the exponential inter-arrival time assumption may need to be rejected due to the re-association events triggered by infrastructure timers, while the independence assumption may not hold due to users arriving in a batch.

Association times are found to be Weibull distributed in [3], in a campus trace from 2003, while *biPareto* distribution is demonstrated already in [7]. Instead, [5] models the association times with order 5 phase-type distribution in a large public hot-spot trace from 2010. In our analysis, only a very few APs showed heavy tail association times, and two-stage hyper-exponential distribution provided an accurate model in most of the cases, reflecting the presence of temporally static users and users on the move [16]. The main reason for non-compliance is the dominance of very short visiting times, most likely due to users' mobility in a dense AP deployment.

VI. Summary and discussion

This paper presents a study of access patterns in a university wireless network, based on a relatively recent and previously unexplored wireless data measurement. We introduce and describe a detailed methodology for processing Eduroam traces, which are readily available in authentication servers in wireless networks supporting Eduroam. The main contributions of our work are the characterization of user arrivals at the network APs and the model of user visiting times. Our findings are partially in accordance with previously established results regarding the observed arrival processes at APs, which often can be modeled with time-varying Poisson process; however, we identify the cases and the reasons when the model is not suitable. With respect to modeling user visiting time, we find that in a large majority of APs, the visiting time distribution can be fitted with two-stage hyper-exponential distributions, reflecting the dominance of mobile and handheld devices in the network.

VII. Acknowledgments

We thank KTHLAN staff members Hans Berggren and Anders Hillbo for providing us with the association records.

This work was partially supported by the European Commission Horizon 2020 Programme RIFE Project Grant No. 644663.

References

[1] F. Ganji, Ł. Budzisz, F. G. Debele, N. Li, M. Meo, M. Ricca, Y. Zhang, and A. Wolisz, "Greening campus WLANs: Energy-relevant usage and mobility patterns," *Elsevier Comput. Netw.*, vol. 78, 2015.

[2] Y. Ding, B. Han, Y. Xiao, P. Hui, A. Srinivasan, M. Kojo, and S. Tarkoma, "Enabling energy-aware collaborative mobile data offloading for smartphones," in *Proc. IEEE SECON*, June 2013.

[3] M. Papadopouli, H. Shen, and M. Spanakis, "Characterizing the duration and association patterns of wireless access in a campus," in *Proc. 11th European Wireless Conference*, Apr. 2005.

[4] M. Papadopouli, H. Shen, and M. Spanakis, "Modeling client arrivals at access points in wireless campus-wide networks," in *14th IEEE Workshop on Local Metropolitan Area Networks*, Sep. 2005.

[5] A. Ghosh, R. Jana, V. Ramaswami, J. Rowland, and N. K. Shankaranarayanan, "Modeling and characterization of large-scale Wi-Fi traffic in public hot-spots," in *Proc. IEEE INFOCOM*, Apr. 2011.

[6] Y.-C. Chen, J. Kurose, and D. Towsley, "A mixed queueing network model of mobility in a campus wireless network," in *Proc. IEEE INFOCOM*, Mar. 2012.

[7] L. Oliveira, K. Obraczka, and A. Rodríguez, "Characterizing user activity in WiFi networks: University campus and urban area case studies," in *Proc. ACM MSWIM*, Nov. 2016.

[8] L. Johansson and S. Winter, "F-Ticks - A Federation Log Format," IETF, Internet-Draft draft-johansson-fticks-00, 2013. [Online]. Available: https://datatracker.ietf.org/doc/html/draft-johansson-fticks-00.

[9] Google Maps API – Distance Matrix API. [Online]. Available: http://developers.google.com/maps/documentation/distance-matrix.

[10] G. Ross, "Sequential change detection in the presence of unknown parameters," *Statistics and Computing*, vol. 24, no. 6, Nov. 2014.

[11] G. Ross, "Parametric and nonparametric sequential change detection in R: The `cpm` package," *J. Stat. Softw*, vol. 66, no. 3, 2015.

[12] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown," *Journal of the American Statistical Association*, vol. 64, no. 325, 1969.

[13] W. Brock, J. Scheinkman, W. D. Dechert, and B. LeBaron, "A test for independence based on the correlation dimension," *Econometric Reviews*, vol. 15, no. 3, 1996.

[14] B. Mizrach, *Using U-Statistics to Detect Business Cycle Nonlinearities*. Dordrecht: Springer Netherlands, 1994.

[15] A. Feldmann and W. Whitt, "Fitting mixtures of exponentials to long-tail distributions to analyze network performance models," in *Proc. IEEE INFOCOM*, Apr. 1997.

[16] U. Kumar, J. Kim, and A. Helmy, "Changing patterns of mobile network (WLAN) usage: Smart-phones vs. laptops," in *Proc. IWCMC*, July 2013.