Study of YouTube Demand Patterns in Mixed Public and Campus WiFi Network

Jukka-Pekka Laulajainen*, Åke Arvidsson†, Timo Ojala‡, Janne Seppänen* and Manxing Du§
*VTT Technical Research Centre of Finland, Oulu, Finland, first.lastname@vtt.fi
†Ericsson Research, Stockholm, Sweden, first.lastname@ericsson.com
‡University of Oulu, Department of CSE, first.lastname@ee.oulu.fi, Oulu, Finland
§Acreo Swedish ICT AB, first.lastname@acreo.se, Stockholm, Sweden

***Abstract*—In this paper we study traffic patterns in a large municipal WiFi network and in particular those of the most bandwidth hungry application, *viz.* YouTube, for which we provide a detailed analysis of demand in different geographical areas and over time. We consider the possibilities to reduce network traffic and increase Quality of Experience (QoE) by serving repeated requests for YouTube videos from caches placed either at the network head end, at the wireless access points, or in the user devices. Our data confirms that a significant part of the YouTube traffic can be served by such devices and that there exists a potential to optimize caching performance by exploiting the content demand locality. We also discover a previously unknown pattern of periodicity in content demand and present a simple example of how to exploit this in cache design.**

***Keywords*—Traffic analysis, wireless networks.**

I. Introduction

According to several recent studies, the use of different services delivered over-the-top is increasing rapidly in both cellular and fixed networks [1], [2]. The largest category in terms of traffic volume is video which thus is the main driver behind the data traffic explosion. The situation is problematic for network owners as the transmission capacity is consumed by applications that only generate small (volume based) revenues. Various ways for optimizing the use of networks are therefore being studied, such as content caching in network elements or terminal devices. Understanding the usage characteristics of networks and applications is critical for the design of such optimization mechanisms.

There are a lot of studies concerning the high-level application usage of wireless and wired networks. In [3], the authors present a thorough analysis of the Google WiFi network deployed in Mountain View, California. Similar to our approach, they consider differences between distinct geographical locations in their paper, but ideas on exploiting the differences in optimization of use of networks are not presented in the paper. Other larger-scale studies of WiFi networks have mostly been conducted in campus networks such as in [4], [5] and [6]. Traffic characteristics of fixed residential broadband networks have been reported *e.g.* in [7] and [8].

Most of the previous works on characterizing the content demand of network applications have concentrated on Web traffic such as [9] and [10]. More recently, the interest has moved to file sharing ([11], [12], [13]) and finally to multimedia applications such as YouTube ([14], [15], [16], [17], [18]). The majority of these focus on the content usage characteristics in wired networks, while the information concerning wireless networks is limited. In [19] and [20], the web browsing patterns of mobile users were analyzed, and the locality of web requests within a WiFi network was studied in [21]. Our paper contributes to this line of work by studying the usage of the most popular individual application, YouTube, in a large public WiFi network. In addition, we split our user base in a unique way between users in the city area, at the university campus, and in the municipal schools. It should be noted, that our focus is on finding out behaviour that could be considered for exploitation while a discussion of the business potential of such exploitation is out of the scope of this paper.

The combination of large scale (1441 access points, 60315 user devices) and versatile profile (outdoor and indoor coverage in city centre, municipal buildings, campuses, and private properties) of the network and the rich data collected over a relatively long period of seven weeks differentiates our paper from previous studies in terms of its experimental setting. Moreover, our findings on differences between usage environments have not been presented in other studies. The main contribution of our paper is, however, the discovery of periodicity of content demand. This has potential to provide several exploitation possibilities in network optimization once studied further.

II. Network and Data

Our analysis is based on measurements performed in a public WiFi network located in the Oulu region in Finland during seven weeks between November 2012 and January 2013. In the following subsections we describe the network and our measurement methodology.

A. The panOULU Network

The panOULU (public access network OULU)[1] network is a regional municipal WiFi network located in the Oulu region in northern Finland. The network is provided jointly by a consortium of municipalities, public institutes in research and education, and private companies. It provides free and open wireless Internet access to anyone in its coverage area. The joint consortium was established in 2003, and the network has been expanding since then to have indoor coverage in basically all public buildings in the city of Oulu, outdoor coverage in downtown Oulu and other selected locations,

[1] http://www.panoulu.net/

such as sport centers. In addition, the network coverage has also expanded to nearby towns around the city and private companies willing to provide wireless connectivity to their customers and visitors [22].

At the moment of this study, the network had 1441 access points (APs) representing a mixture of IEEE 802.11a/b/g/n technologies and manufactured by Cisco, Siemens, Strix and Linksys. Providers have recently been updating their WiFi zones from stand-alone 802.11a/b/g configuration to controller based 802.11n configuration. While downtown Oulu and a few other areas are blanketed with Strix OWS series multi-radio mesh APs, coverage in other parts is provided in a hot spot manner. All APs advertise the same SSID ("panoulu") such that they appear to be part of one large wireless network from the users' point of view. The hot spots and mesh root APs are connected to an Ethernet-based aggregation network with fixed xDSL or Ethernet links. The network is connected to the public Internet via a gateway provided by a local ISP and there are no limitations or restrictions on the use of the network.

B. Data Collection

We collected three principal types of data for the purposes of this study, overall application statistics, YouTube packet traces and session statistics, over a period of 49 days from November 2012 to January 2013.

First, the overall statistics of application usage in the network were gathered by connecting a commercial PacketLogic PL8720 Advanced DPI (deep packet inspection) system from Procera Networks [2] to the Internet gateway. Although the DPI system is designed for bandwidth management and policy enforcement, we used it as a passive measurement probe. The system uses a proprietary traffic identification engine to generate statistics about network usage and to accurately filter network traffic in real time. The system was configured to calculate statistics of application usage in terms of bandwidth consumption. According to the statistics, 32% of the bytes transmitted in the network were generated by streaming media applications. YouTube was by far the most popular individual application in this category, hence we focused in YouTube traffic characteristics in this work.

Second, the necessary parts of all YouTube flows were collected with the same commercial tool as the application usage statistics, but with another parallel configuration. For the purposes of our analysis, we configured the system to capture the raw bytes belonging to the beginnings of all YouTube sessions into *pcap* files and to store the result on a disk.

Third, a centralized server logs different sessions in terms of user device id, start and end times, and AP ids. As the panOULU consortium is based on voluntary participation, it does not dictate how its members should implement their WiFi zones and what kind of usage data they are required to report. Consequently, session data is only available for a subset of all APs. In this study we use session statistics of about 5 000 000 sessions recorded during the measurement period at three different categories of APs (geographical areas) totalling 606 APs:

- **Campus:** 110 APs located at the main campus of the University of Oulu 5 kms north of downtown Oulu.
- **City:** 98 wireless mesh APs providing outdoor coverage in downtown Oulu and 289 APs in public municipal buildings around the city (excluding schools).
- **School:** 109 APs located in schools around the city of Oulu. The schools include elementary schools, with pupils aged between 6 and 15 years, and high schools, with students aged between 15 and 18 years.

III. YouTube Usage

Our analysis of YouTube usage is based on capturing the beginnings of all YouTube flows on a disk and correlating these captures offline with the session database. User privacy is protected in the process by using hashed identifiers. We are able to bind video views to individual APs for all sessions in the database. In more detail, we use the AP the user device was associated to when the video was requested and, although we note that user movements or varying coverage may have resulted in one or more changes of APs before the video was completed (or interrupted), we also note that YouTube video clips (and viewing times) are typically quite short hence we assume that this does not have a major impact on our results.

Moreover, in addition to the AP, we recorded the YouTube content identifiers and the user agent information, which were available for all videos viewed during the measurement period. In more detail, we examined HTTP GET requests for media content to YouTube and identified videos from the 16 hexadecimal digit identifiers in the "videoplayback" field and terminal types from the text values in the "User-Agent" field. Altogether our measurement covered over 200,000 video views out of which we were able to locate roughly 50,000 with city APs, 20,000 with campus APs and 20,000 with school APs. The rest of the videos were viewed with APs not connected to the session database (due to technical or administrative reasons), so they are not included in the results comparing different usage environments.

We start our analysis by looking at the temporal aspects. Figure 1 shows how the YouTube views are distributed over the different days of the week. As can be expected, the number of views at campus APs and school APs is the lowest during the weekend when there are no classes and the highest on Mondays when weekend findings are shared between friends. In the city APs, this effect is not visible but the differences between the number of views between different days are much lower, although we note that the number of views is the highest on Fridays and the lowest on Sundays.

We remark that these variations can be caused by (i) varying number of requests per user, (ii) varying number of requesting users or (iii) both. A further investigation shows that, while the average number of views per active user seen requesting at least one video during a day is around 6 in all AP categories during the weekdays, this number does in fact increase for campus and school APs during weekends but is almost constant in the city, cf. the dots in Figure 1. On the other hand the number of users (Table I) again is almost constant during week days but does in fact decrease for campus and school APs during weekends. Thus we draw two conlusions:

[2] http://www.proceranetworks.com/

TABLE I. FRACTION OF USERS PER DAY OF WEEK

AP category	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
Campus	0.22	0.18	0.20	0.16	0.17	0.04	0.03
City	0.16	0.15	0.15	0.14	0.17	0.13	0.10
School	0.22	0.20	0.20	0.16	0.18	0.02	0.02

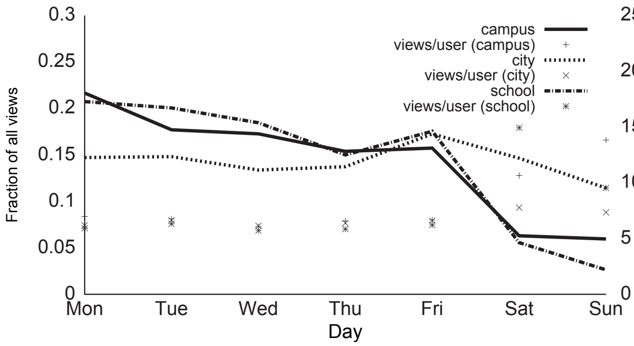

Fig. 1. YouTube use, days.

first, the relatively constant demand at city APs is a result of relatively constant number of users and relatively constant demand per user; second, the variations in demand at school APs and campus APs are a result of a noticeable increase in demand per users and an even more pronounced decrease in the number of users.

While campus and school usage resemble each other with respect to week days, there is a significant difference between them with respect to peak hours (Fig. 2); the campus peak occurs between 11 *a.m.* and 4 *p.m.* while the school peak is between 9 *a.m.* and 2 *p.m*. For the city we note that the lower variation with respect to week days also applies to variation with respect to hours with a less pronounced but extended peak period between 12 *a.m.* and 7 *p.m.* in the city. The average amount of video views for each user seen requesting at least one video during an hour is around 4 for all categories during daytime while average nighttime users are more active, in particular those at APs in schools or at the campus.

As our YouTube traces include all data from the beginning of each YouTube flow, we are able to record the user agent string from the HTTP request. Using the HTTP user agent information, we categorize the user devices as either PCs or mobile devices. The data also includes a few other devices such as TV sets and game consoles, but we have excluded them from the analysis because their share of the total traffic is insignificant. Figure 3 shows how YouTube usage varies over the week for both device categories. PC usage seems to be more stable than mobile usage and for the latter we note that weekday usage is about double the weekend usage. Similarly, Figure 4 shows how YouTube usage varies over the day for both device categories. Mobile usage peaks earlier in the day than PC usage, and a deeper analysis shows that this is due to heavy usage of mobile devices in schools during school days and school hours, see Figures 5 and 6, while no such peaks are seen for APs in the city or on the campus. A general observation (not shown) is that PCs dominate during the night hours in all three environments.

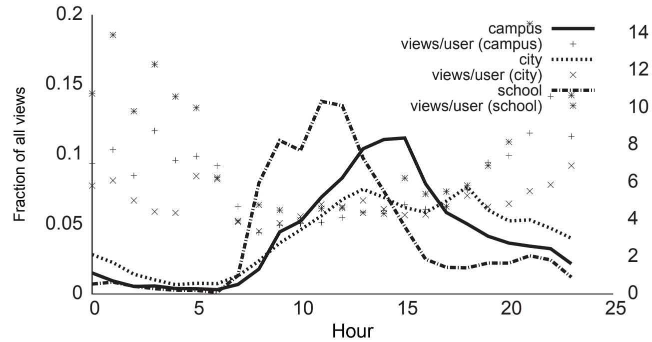

Fig. 2. YouTube use, hours.

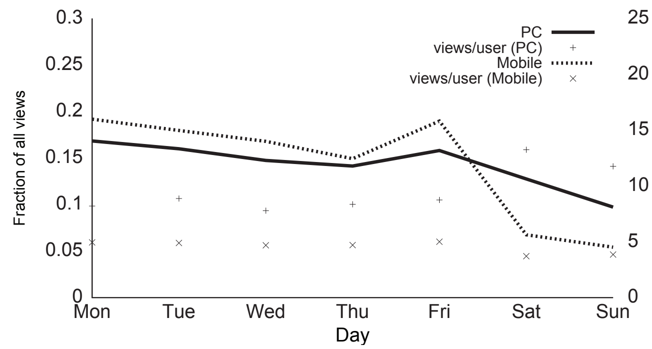

Fig. 3. YouTube use, device types, days.

IV. POTENTIAL TO SERVE YOUTUBE REQUESTS BY CACHING

We commence our exploration of the possibilities of serving YouTube requests via caching by analysing content popularity in the YouTube traffic. At this stage we look at the statistics of the whole data but we shall include finite cache size effects later in this section.

Plotted in log-log scale, the number of requests for the most popular items ordered according to their popularity, the Figure 7 resembles a Zipf distribution, which implies that a finite sized cache could serve a significant amount of YouTube requests. This observation is well in line with literature (*e.g.* [15]).

To further assess the potential gain from caching, we plot the amount of repeated requests for the same content in Figure 8. When considering all YouTube video views in the network, 27% of the video files were viewed at least twice and the most popular video was viewed 513 times. Within the smaller categories of campus, city and school, the amount of videos viewed only once increases, but even in the campus APs, with the lowest similarity between the requests, more than 18% of videos were viewed at least twice.

Next we consider the level of individual user devices. We define a user device as a unique pair of a MAC address and a user-agent which corresponds to a natural position for a possible content cache. Figure 9 shows the CDF of repeated views of a particular YouTube video on a particular device. We see that over 16% of all requests are such "double

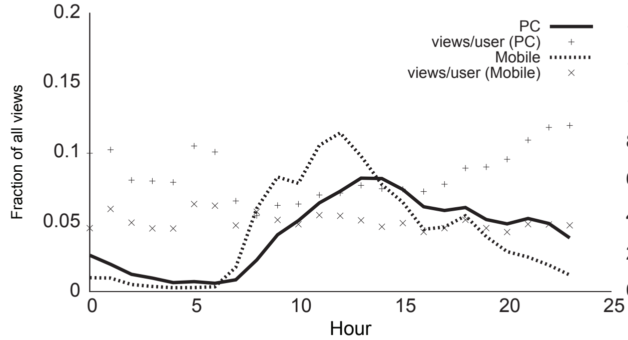

Fig. 4. YouTube use, device types, hours.

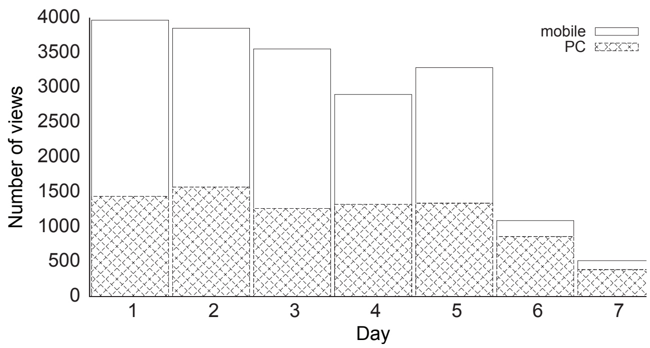

Fig. 5. YouTube use, days and device types, school.

repeats" (same user and same content). Also in this case, restricting the examination to the separate AP categories yields smaller amounts of repeated requests (since only repeats in the same AP category count) and the order of AP categories is interestingly the same as previously. The most repetitive requests are made by users connected to city APs and the least repetitive ones by those connected to campus APs.

We define three measures of content cacheability, *viz.* the total hit rate h_{T} (the fraction of requests that could have been served from a cache), the local hit rate h_{L} (the fraction of intra user hits, *i.e.* hits that could have been realised by local caches inside the devices) and the global hit rate h_{G} (the fraction of inter user hits, *i.e.* hits that must be realised by global caches outside the devices). Note that the total hit rate thus is the aggregate of local hits (intra user hits that could have been made *inside* the devices) and global hits (inter user hits that must be made *between* the devices). In formal notation the total hit rate is defined as

$$h_{\mathrm{T}} = 1 - \frac{U}{T}, \tag{1}$$

where U and T are the number of unique video requests and the total number of video requests respectively, the local hit rate is defined as

$$h_{\mathrm{L}} = 1 - \frac{\sum_{\forall u} U_u}{T}, \tag{2}$$

where U_u is the number of unique video requests from user

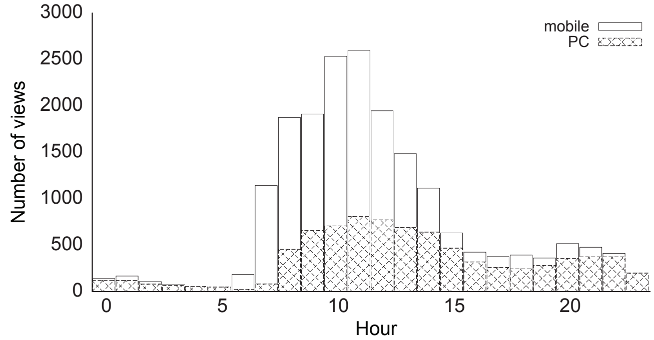

Fig. 6. YouTube use, hours and device types, school.

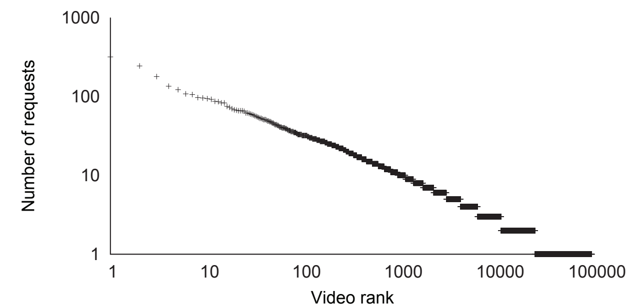

Fig. 7. YouTube use, number of requests per video.

u, and the global hit rate is defined as

$$h_{\mathrm{G}} = 1 - \frac{\sum_{\forall u} U_u - U}{T}. \tag{3}$$

These measures count the cumulative hit rate starting from an empty cache. To assess the cache performance in normal operative state, we also calculated the total hit rate for the last week of measurements as

$$h_{\mathrm{T}'} = 1 - \frac{U'}{T'}, \tag{4}$$

where U' is the number of unique video requests during the last week of measurements, *i.e.* the number of videos never seen in our trace before the last week of measurements, and T' is the total number of video requests during that week.

Table II shows the caching gains calculated over all our data, over user device categories, and over AP categories.

The results show that, although larger groups predictably provide higher hit rates, small but in some sense homogeneous groups can also provide high hit rates, cf. the school category which despite having the smallest number of users exhibits the highest hit rate. One possible explanation in this case is

TABLE II. CACHING GAIN POTENTIAL

Category	Users	Videos	Requests	$h_{\mathrm{T}} (h_{\mathrm{T}'})$	h_{L}	h_{G}
All	11893	117912	207099	0.43 (0.48)	0.23	0.20
PC	6735	92705	157154	0.41 (0.46)	0.23	0.18
Mobile	4896	29606	47069	0.37 (0.42)	0.24	0.13
City	4489	33431	49378	0.32 (0.37)	0.21	0.11
Campus	1774	14440	19954	0.28 (0.32)	0.22	0.06
School	1711	13196	19589	0.33 (0.37)	0.22	0.11

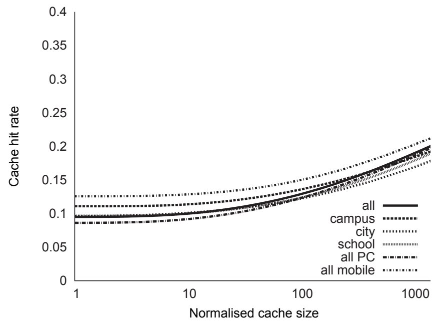

Fig. 10. Hit rate for LRU caches in terminals.

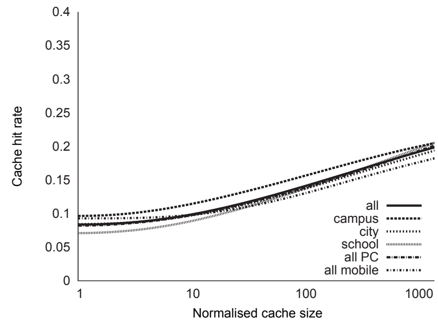

Fig. 11. Hit rate for LRU caches in APs.

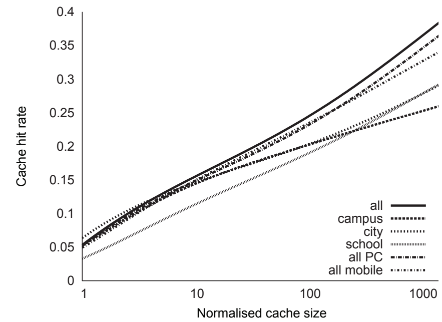

Fig. 12. Hit rate for LRU cache at the Internet edge.

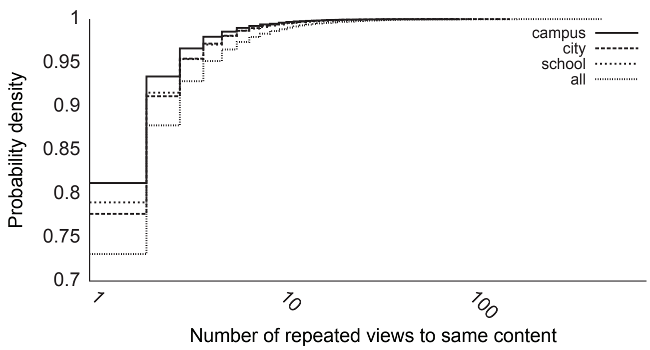

Fig. 8. YouTube use, CDF of requests per video.

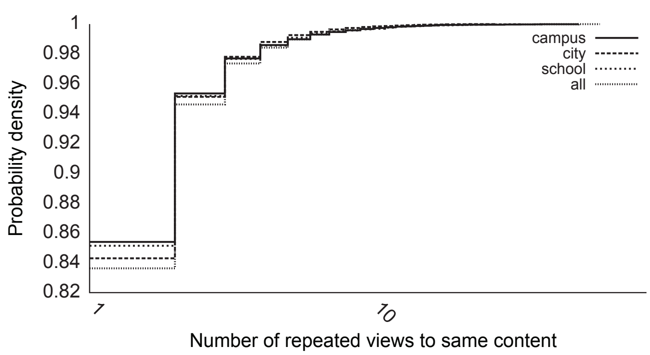

Fig. 9. YouTube use, CDF of requests per video per user.

that school kids may be more likely to replay videos to their friends.

Above, our analysis was based on an assumption of an infinite size cache. For more realistic results, we ran the recorded YouTube requests through (simulated) simple LRU (Least Recently Used) caches, all of which are assumed to store a defined number of video clips and discard the least recently used clip when a new clip is stored and the cache is full. Figures 10, 11 and 12 show the results in terms of traffic reduction by deploying these caches either at terminals, at WiFi APs, or at the Internet edge, respectively. Cache sizes are expressed as percentages of the number of unique video clips per day and unit at which the cache resides in order to visualize differences related to specifics of separate categories rather than the sizes of these categories. Thus, *e.g.*, cache size 10 in a terminal corresponds to a cache that can store 10% of the average number of unique video clips requested per day by a terminal, and cache size 1000 in an AP corresponds to a cache that can store 1000% of the average number of unique video clips requested per day at an AP.

Overall, by adding a sufficiently large YouTube cache into the network, 48% of the video requests could be served locally and, similarly, by adding sufficiently large YouTube caches to the user devices, more than 23% of requests could be served without using the network at all. The results thus show that simple LRU caches with a storage capacity equivalent to the number of unique video clips consumed in ten days will be able to deliver hit rates close to those calculated for infinite caches. This result is well in line with those presented in [14]. When comparing the results between different categories in Table II, we see that interestingly the global hit rate of PC users is much higher than that of mobile users, implying that there are more similarities between PC users as a group than between mobile users as a group. On the other hand, the requests of mobile users seem to be more self-similar, favouring deploying caches in mobile devices. This is also visible in Figure 10. With respect to AP categories, the best performance is obtained for the campus category (Fig. 11) while the performance is relatively poor compared to the other categories for a cache at the Internet edge (Fig. 12). This suggests that the consumed content differs a lot around the campus area, but there is a noticeable similarity between the users sharing the same AP. Further, we note that the cacheability results for the school category grow almost linearly with the size of the cache, what we interpret as a sign of homogeneous content consumption among the school kids.

After observing spatial demand locality, we investigate temporal demand characteristics. We start with an assumption that video requests during certain hours may have more in common with requests during the same hours (but possibly different days) than with requests during other hours. To test this demand locality in time, we take random groups of 20,000 video requests from hour-of-day based subgroups of all requests, run them through our simulated LRU cache, and compare the results with similar sized random groups taken from all hours of the day. Figure 13 shows the cache hit ratios. All time restricted subgroups show higher hit ratio than the results for all data. We take this as a sign of locality of demand not only in space but also in time.

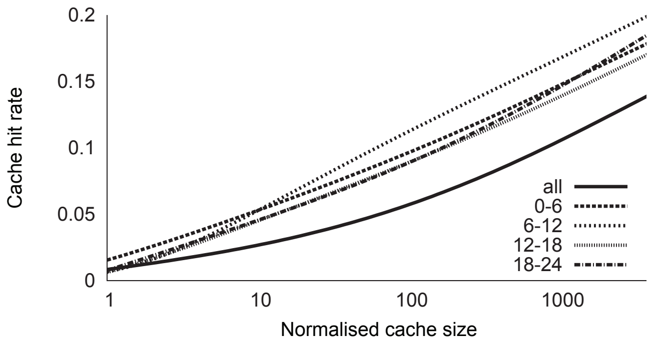

Fig. 13. Hit rate, hour-based groups

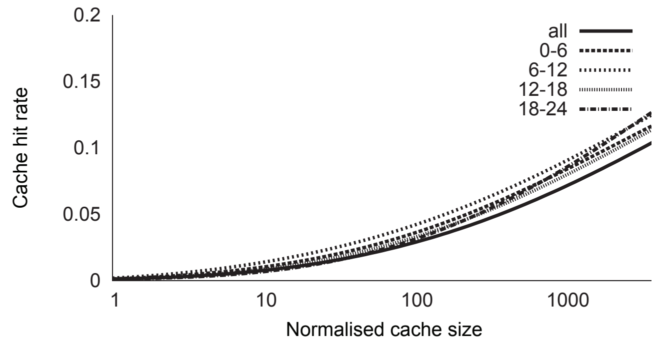

Fig. 14. Hit rate, hour-based groups, modified

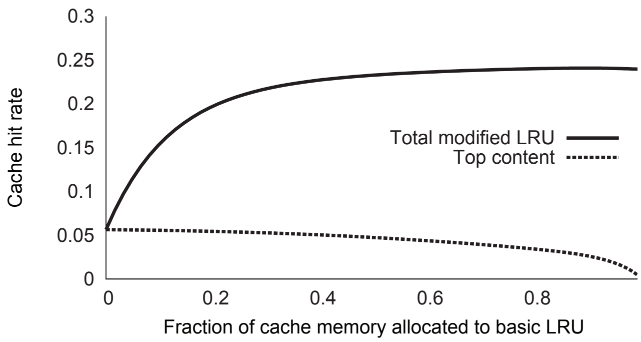

Fig. 15. Hit rate for modified LRU cache

To understand this further, we modify our test by removing same day same user repeated requests to focus on the periodicity of demand between days. Figure 14 presents the results for the modified case. It is seen that the differences in cache hit ratio between the time-of-day based subgroups and the whole-day data are smaller than before, but still we see a benefit from splitting the cache in time and thus the locality in time is not only due to bursts of repeats from single users but from views by different users at the same time and/or by the same user or users at different days. We assume that the periodicity of the demand may be at least partially caused by users having regular active periods which differ between users.

To examine this assumption, we define for each six hour time period p, (*e.g.* the hours between 6 *a.m.* and 12 *p.m.* and so on) a reference group of users G_p who were active during that period in the first week of our measurements and another, reference group of the same number of randomly selected users $\bar{G}_p$ who were active outside period p during the same period. For each period p we then compute the fraction of users that are seen during p in over the remaining six weeks of our measurement. The results are shown in Table III and we see that, *e.g.*, users who were active in the night (0–6) during the first week are more likely (46%) to be active in the night during the remaining weeks than random users seen at other times (4%). A similar difference can be seen for evenings (18–24) while this pattern is less evident for mornings (6–12) and afternoons (12–18). It is thus clear that the periodic content demand to some extent indeed can be explained by

TABLE III. User periodicity

Group	0–6	6–12	12–18	18–24
G_p	0.46	0.06	0.04	0.13
$\bar{G}_p$	0.04	0.04	0.04	0.04

periodic user activity. We conclude that, with proper ways of distributing different types of content between different physical servers, it should be possible to save power by running a smaller number of dedicated servers during nights without sacrificing hit rates.

Finally, we demonstrate the exploitation of content demand periodicity with a simple example. We modified our LRU cache implementation by reserving parts of the memory for the most popular content for that particular time of day. The hours of a day are split into four six hour periods, and the most popular video clips for each period are copied to the cache at the beginning of the period (top videos of other periods are stored outside of the cache). When video clips are being requested from the cache, the memory containing the top videos from the previous day's corresponding period is checked first and if there is no match, the LRU part of the cache is used in the normal way. Figure 15 presents the performance of this modified LRU cache. For simplicity, we fixed the total size of the cache to the amount of unique video clips requested per day on average during our measurement period, and the figure can be compared to Figure 12 with the basic LRU operation. The results show that with optimal memory allocation between time-based top cache and basic LRU memory, an increase in cache hit rate can be achieved. The maximum hit rate of 24.1% was achieved with 87.5% of memory allocated to basic LRU. This improvement of less than 1% in hit rate in comparison to allocating the whole cache to basic LRU can be considered minor, but with a more sophisticated way of using the periodicity of demand, higher benefit could potentially be obtained. Moreover, the hit rate is typically not the only parameter to look at in cache design. This kind of modified LRU could be useful for example if the cache had separate fast and slow memory, and the top content of the hour could be served faster to provide better QoE.

V. Conclusions

In this paper we have studied YouTube usage in a large municipal WiFi network. During our seven week data collection period, we collected information related to more than 200,000 video views using over 600 separate WiFi APs. We presented the daily and weekly YouTube usage patterns for three different geographical areas and looked at the differences in YouTube usage between different user device types. The variance of the amount of YouTube use was much lower on both daily and weekly level in the city APs than in campus and school APs in which the user groups and their schedules are more uniform. Similarly, PC usage was seen to be more stable than mobile use, and the night-time users on average used YouTube heavier than daytime users.

We also looked at the similarity between YouTube requests and showed that there is a significant potential for saving bandwidth and improving QoE by caching YouTube content either at the network head end, at the network access points or in the user devices. The caching gain potential varies between the AP categories and according to our results the requests of city and school users have bigger potential for device caching than the requests of university campus users. Understanding the differences in content demand in different environments could be used to increase caching performance when deploying local caches.

Finally, as a novel contribution, we discovered periodicity in the content demand and presented a first, simple example of exploiting it in cache design. Our findings on content demand periodicity should be confirmed by further research and ways of exploiting it should be studied. In this paper our focus was only in YouTube, but the caching potential and content demand characteristics of other popular applications should also be studied in the future.

REFERENCES

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2012–2017," 2013.

[2] "Ericsson mobility report: On the pulse of the networked society," 2013.

[3] M. Afanasyev, T. Chen, G. M. Voelker, and A. C. Snoeren, "Analysis of a mixed-use urban wifi network: when metropolitan becomes neapolitan," in *ACM SIGCOMM conference on Internet measurement*. ACM, 2008, pp. 85–98.

[4] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *MobiCom*. ACM, 2004.

[5] D. Kotz and K. Essien, "Analysis of a campus-wide wireless network," in *MobiCom*. ACM, 2002.

[6] A. Gember, A. Anand, and A. Akella, "A comparative study of handheld and non-handheld traffic in campus wi-fi networks," in *PAM*, 2011, pp. 173–183.

[7] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband internet traffic," in *IMC*. ACM, 2009, pp. 90–102.

[8] M. Dischinger, A. Haeberlen, K. P. Gummadi, and S. Saroiu, "Characterizing residential broadband networks," in *IMC*, vol. 24, no. 26, 2007, pp. 43–56.

[9] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," in *INFOCOM'99*. IEEE, 1999.

[10] C. R. Cunha, A. Bestavros, and M. E. Crovella, "Characteristics of www client-based traces," Boston University Computer Science Department, Tech. Rep., 1995.

[11] S. Sen and J. Wang, "Analyzing peer-to-peer traffic across large networks," *IEEE/ACM Transactions on Networking (ToN)*, vol. 12, no. 2, pp. 219–232, 2004.

[12] A. Wierzbicki, N. Leibowitz, M. Ripeanu, and R. Wozniak, "Cache replacement policies revisited: The case of p2p traffic," in *CCGrid*. IEEE, 2004, pp. 182–189.

[13] B. Ager, F. Schneider, J. Kim, and A. Feldmann, "Revisiting cacheability in times of user generated content," in *INFOCOM Workshops*. IEEE, 2010, pp. 1–6.

[14] Å. Arvidsson, M. Du, A. Aurelius, and M. Kihl, "Analysis of user demand patterns and locality for youtube traffic," in *ITC*. IEEE, 2013.

[15] F. Guillemin, B. Kauffmann, S. Moteau, and A. Simonian, "Experimental analysis of caching efficiency for youtube traffic in an isp network," in *ITC*, 2013.

[16] J. Erman, A. Gerber, K. K. Ramadrishnan, S. Sen, and O. Spatscheck, "Over the top video: the gorilla in cellular networks," in *IMC*. ACM, 2011, pp. 127–136.

[17] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in *IMC*. ACM, 2007, pp. 15–28.

[18] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Watch global, cache local: Youtube network traffic at a campus network: measurements and implications," in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 681 805–681 805.

[19] F. Qian, K. S. Quah, J. Huang, J. Erman, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Web caching on smartphones: ideal vs. reality," in *MobiSys*. ACM, 2012, pp. 127–140.

[20] A. Adya, P. Bahl, and L. Qiu, "Analyzing the browse patterns of mobile clients," in *ACM SIGCOMM Workshop on Internet Measurement*. ACM, 2001, pp. 189–194.

[21] F. Chinchilla, M. Lindsey, and M. Papadopouli, "Analysis of wireless information locality and association patterns in a campus," in *INFOCOM 2004*. IEEE, 2004.

[22] T. Ojala, J. Orajärvi, K. Puhakka, I. Heikkinen, and J. Heikka, "panoulu: Triple helix driven municipal wireless network providing open and free internet access," in *C&T*. ACM, 2011, pp. 118–127.