A Spatio-Temporal Analysis of Mobile Internet Traffic in Public Transportation Systems

A View of Web Browsing from The Bus

Fangzhou Jiang*†, Kanchana Thilakarathna*, Mohamed Ali Kaafar*, Filip Rosenbaum‡, Aruna Seneviratne*†

*National ICT Australia, †The University of New South Wales, Australia, ‡Oxyfi, Sweden

ABSTRACT

Mobile devices are ubiquitous, and mobile-generated traffic is arguably a major component of today's web traffic. In particular, the use of smart-phones whilst commuting using public transport is a very popular and common practice in many countries. Mobile commuters however, often suffer from poor performance due to limited bandwidth and/or intermittent network coverage. This paper provides insights into the characteristics of web traffic generated by mobile commuters in these challenged conditions of public transportation systems.

We use a dataset collected from 22 Inter-city Buses running on 6 different routes over 5 weeks in Sweden. By analyzing content similarity in time and across different routes, we discover a number of findings that reveal the existence of a spatio-temporal correlation of content popularity and that shed light on diurnal patterns of behavior of mobile commuters. We study popular content accessed by commuters and show that *Social Networking* and *News* content are predominant and are mutually exclusive. One of the salient findings is that mobile users' interest on buses is very concentrated, with 35% of the popular content solely accessed on a single day during the 5 weeks, and more than 70% of the popular content from a given day is accessed during one single hour of the day. We also observe high content similarity between specific routes which suggests that content caching within the bus can significantly improve user web experience. Our results indicate that based on the observed strong spatio-temporal correlation of content requests of mobile commuters, caching content inside the buses leads to a daily hit rate ranging from 10 to 20%, with a 20% savings of the daily bandwidth usage.

Categories and Subject Descriptors

C.2.1 [**Computer Communication Networks**]: Wireless communication; C.2.3 [**Computer Communication Networks**]: Network monitoring

CHANTS'15, September 11, 2015, Paris, France.
 ISBN 978-1-4503-3543-0/15/09 ...$15.00.
DOI: http://dx.doi.org/10.1145/2799371.2799379.

Keywords

Mobile Content Distribution, Spatial and Temporal Correlation, Public Transport

1. INTRODUCTION

Mobile data traffic has already surpassed the total wired Internet traffic and it is expected to grow exponentially [4]. An increasing portion of this traffic is due to the use of smart mobile devices. It has been observed that 83% of commuters use their smartphone while on-board trains and buses in the US [9]. As a result, many companies around the world have been enabling cheaper network services on public transport systems [15, 14]. Moreover, operators are struggling to adapt to the higher traffic levels and to provide satisfactory Quality of Experience(QoE) to users as the capacity bottleneck is typically at the edge of the network [7]. It is even more challenging when the last hop of the network is connected to a mobile vehicle such as a train or a bus. Therefore, traditional content delivery mechanisms, such as using caches within the network is not effective in the context of public transportation systems.

We postulate that commuters traveling on a public transportation system, at given time, i.e. the same bus, are likely to be interested in similar content. For example, students traveling to a university would have different interests to commuters traveling to a city's financial district. In addition, regular commuters are likely to access similar content periodically, i.e. on a daily basis. Previous work studied content popularity characteristics in cellular and device-to-device networks [5, 16]. Likewise, a few others studied content co-location in public transport [13] and mobile content access characteristics at a base-station level [7, 17]. However, little is known about spatial and temporal characteristics of mobile commuters, in particular when accessing Internet content while "on the move".

To test out this hypothesis, and to gain insights into ways of providing effective content delivery services in public transportation systems, we analyze a dataset collected from an Internet service provider in public transport buses in Sweden. We focus our analysis on spatial and temporal patterns of similarity and diversity of content requests that may be used for the design of novel content delivery architectures for transport networks. We observe the following:

- *Social networking* and *News* services are the predominant services, and their popularity seem to be mutually exclusive. Commuters tend to access *News* services during their morning commutes and *Social Network* services during the evening commutes.

- Daily similarity of content requests decreases with time while there is a periodic pattern of high similarity between the same days of the week. The cohort of commuters on weekdays are significantly different from the commuters on weekend, which is reflected in the similarity in content requests.

- Interests of commuters are highly concentrated in time and space. Approximately, 35% of popular content accessed during one day of the observation period of five weeks, more than 70% were only requested within one hour. Further, the commuters traveling in the same route have significantly higher similarity of interests compared to inter-route commuters.

- Due to the observed strong spatio-temporal correlation of content requests of mobile commuters, the caching content inside the transport networks, i.e. within the bus, provides considerably higher cache hit rates.

The remainder of the paper is organized as follows: Section 2 presents the related work followed by the details and characteristics of the dataset in Section 3. We then analyze the temporal, spatial and spatio-temporal correlation of content requests in Section 4. In Section 5, the potential of content caching in buses is investigated. Finally, Section 6 concludes the paper.

2. RELATED WORK

Finamore et al. [7] analyses spatial and temporal correlation of content access per-city and per-tower basis, and its impact to proactive caching with pre-loading bundles of content. The results show that the per-city bundle has a better caching performance. However, they only consider a bundling time window of 1 hour and do not to conduct a finer-grained sensitivity analysis. We argue that spatial content access patterns shall be analyzed in a finer granularity than cellular towers. Finer granularities levels of locations might then indeed show similar interests across mobile users. Our paper is, to the best of our knowledge, the first study that investigates spatial and temporal correlation of web browsing activities in the context of small and highly mobile geographical space.

On the other hand, there have been many studies that analyze spatial and temporal content access patterns. In [10], authors observe strong daily and weekly correlation of video request patterns. Jin et al. [8] also discuss long-term and short-term temporal correlation of web requests. Geographic popularity patterns have been further studied in [11], where mobile users show a stronger temporal locality characteristics than traditional requests. Moreover, in [17], authors present the existence of spatial locality among base stations in order to optimize the content placement and efficient content delivery solutions. However, there has been very limited work that investigate the spatio-temporal correlation in a finer granularity in space. In this paper, we take mobile commuters in public transport buses as an example to give a first in its kind analysis of finer-grain spatio-temporal correlation of web browsing.

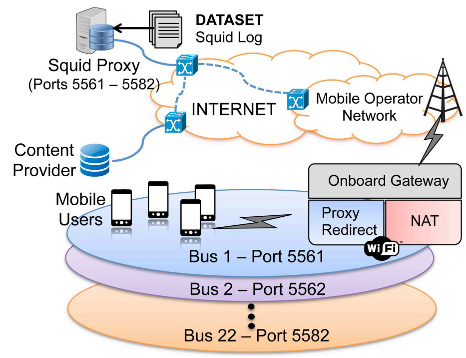
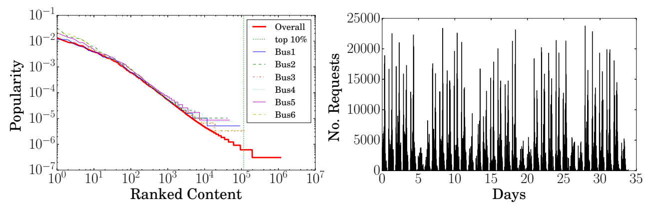


Figure 1: Data collection methodology and network architecture.



(a) Content popularity (b) Workload pattern

Figure 3: Zipf-like popularity distribution and periodic workload pattern

3. WEB BROWSING FROM A BUS

3.1 Dataset

Data collection. We collected a dataset with information of transparent internet content requests from commuter's mobile devices in public buses in Sweden. The data was collected using the system shown in Figure 1. An Onboard gateway device on each bus provides internet access to the passengers via a local *WiFi* network, whilst transparently redirecting all traffic on-wards to a land based *Squid* proxy via a cellular network. All packets from a particular bus are addressed to a unique port in the range of 5561–5582. The proxy responds to all requests by fetching the content either from the local Squid cache or the corresponding content provider. We collected the Squid log for 5 weeks from February to March in 2015. The dataset contains filtered HTTP header information of passenger mobile devices from 22 inter-city buses on 6 different routes. Figure 2 shows the geographical locations of the routes and Table 1 summarizes the basic statistics of the dataset.

Data filtration. We considered only content requests with successful HTTP response (200 OK) code, which accounts for approximately 83% of all requests. The request URL is used to uniquely identify content by taking domain, sub-domain and path of the URL and removing the user specific parameters. For instance, after removing the parameters in `www.skolverket.se/favicon.ico;jsessionid=1742222CE5204EB8D378365819506BEF`, we regard `www.skolverket.se/favicon.ico` as a unique content.

Similar to previous studies [1, 2, 3], the content popularity displays a heavy tail. Figure 3a shows the zipf-like popularity distribution of all content as well as the most "popular" bus in each route (linear log-log plot). The results show that

Table 1: Summary of the dataset

Route #	City A(Population)	City B(Population)	# Buses	# Requests	# Requests/Bus 1	2	3	4	Distance
1	Stockholm(789k)	Södertälje(65k)	4	892,128	222,827	294,998	181,039	193,265	36km
2	Nyköping(30k)	Eskilstuna(65k)	4	505,234	119,838	138,432	53,962	193,002	84km
3	Örebro(130k)	Karlstad(87k)	4	300,657	90,044	45,556	68,549	96,508	111km
4	Karlstad(87k)	Karlskoga(27k)	4	988,333	304,680	211,437	241,787	230,429	67km
5	Karlstad(87k)	Kristinehamn(18k)	3	343,414	111,097	103,294	129,023		43km
6	Karlstad(87k)	Töcksfors(1k)	3	237,485	115,383	32,976	89,126		120km

Figure 2: Geographical map of the bus routes.

Table 2: Distribution of mobile device platforms

Type	Percentage of Requests
Android	42.97%
iOS	38.31%
Windows Desktop	5.34%
Mac OS X	2.19%
Windows Phone	0.94%
Other	10.25%

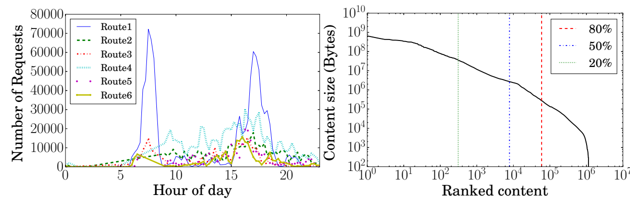
even at the edge of the network, the content popularity still follows a zipf-like distribution. In addition, 30% of URLs are only requested once. Therefore, to remove the long tail part from our analysis, we only consider the top 10% of most popular content requests in the remainder of this paper.

3.2 Web browsing characteristics of commuters

Overall, there are 1,140,757 unique content requests from 26,699 unique domain names. The most pervasive mobile device platforms Android and iOS contributes to more than 81% of the HTTP requests during the 5 week period. Safari browser accounts for 86% of the iOS traffic. However, majority of the Android traffic is divided between the Chrome Mobile (55%) and the Android Browser (33%). The distribution of user mobile device platforms is shown in Table 2.

Figure 3b shows a periodic weekly content request pattern for the 5 weeks. It clearly shows, as expected, that the traffic volumes decrease dramatically during weekends. The aggregated hourly request patterns in Figure 4a shows a noticeable morning and afternoon peak at around 7am and 4pm for some routes such as R1. This is most likely due to the fact that R1 is used by the residents of Södertälje who work in the Stockholm city. Despite the higher temporal peaks in R1, R4 generates a larger number of requests that are distributed through out the day. This reflects the difference of behavioral characteristics of commuters in different cities and routes.

Content-Length field of HTTP header is used as a measure of content size. In Figure 4b, content is ranked according to *no. of requests* × *size* (log-log scale). It can be observed that 309 content items contribute to 20% of total traffic and 59,551, approximately 5% of all content items, accounts for 80% of total traffic.



(a) Daily number of requests in different routes (b) Content size

Figure 4: Diurnal patterns of content requests and content size distribution

Table 3: Distribution of content types

Type	Per. of Requests	Per. of Size
Images	48.9%	26.8%
Text	28.6%	13.4%
Application	18.7%	28.3%
Video	1.2%	22.6%
Audio	0.3%	8.5%
Other	2.3%	0.3%

MIME-Type of HTTP response header is used to categorize content types as tabulated in Table 3. Images account for almost half of the number of requests, followed by text and application. Although videos only account for 1.2% of the requests, they accounts for significant portion, 22.6%, of the traffic volume.

Content Category. We categorized the unique content we identified using McAfee's URL ticketing system [12], and the top 15 categories in terms of total traffic contribution are shown in Table 4. For example, the most popular local news sites are `Aftonbladet`, `Dn` and `Expressen`. In addition, `Instagram`, `Facebook` and `VK` are the among the most popular social networking domains. The ratio of number of requests over number of unique content indicates the average number of requests for each content in the different categories. For example, on average each content in `Web Ads` is requested 18 times, which depicts a large potential for caching.

As shown in Figure 5, *Social Networking* and *News* services are the predominant categories and heavily accessed by commuters. However, the interest of these two categories

Table 4: Top 15 content categories according to domain names

Category	No. Contents	No. Req	$\frac{\|Req\|}{\|Contents\|}$
Internet Services	58,966	464,908	7.9
Social Networking	309,853	391,477	1.3
Uncategorized	30,289	333,763	11.0
General News	96,661	286,637	3.0
Web Ads	14,980	269,790	18.0
Content Server	115,803	233,617	2.0
BizSoftware/Hardware	40,038	189,849	4.7
Business	27,929	129,007	4.6
Marketing/Merch	51,973	92,399	1.8
Entertainment	25,678	65,576	2.6
Software/Hardware	11,384	49,647	4.4
Online Shopping	28,793	41,602	15.3
Games	12,562	41,385	3.3
Sports	18,554	38,455	2.1
Search Engines	4075	36,625	9.0

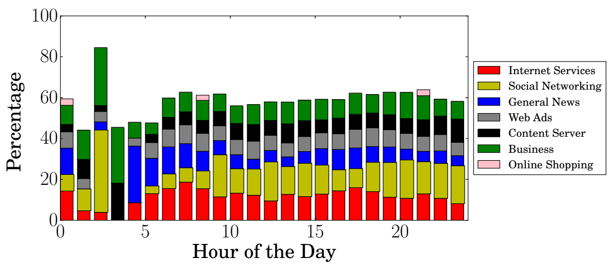


Figure 5: Daily content category distributions

seem to be mutually exclusive. We also observe that commuters tend to access News services during their morning commutes and Social Networking services during the evening commute.
Takeaway: *Content access patterns are heavy tailed even at the edge of the network in terms of both popularity and size. Social networking and News content are predominant and mutually exclusive in popularity. Despite the low mobile video usage by commuters, mobile video accounts for approximately a quarter of total traffic.*

4. SPATIAL AND TEMPORAL ANALYSIS

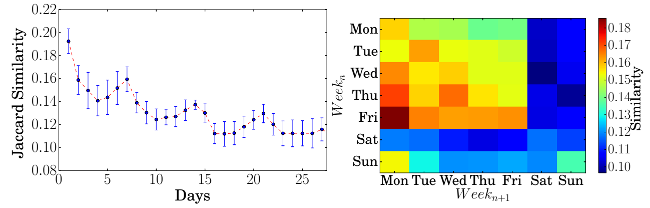
We investigate the existence of spatial and temporal correlation of content requests patterns of commuters in this challenged public transportation systems networks.

4.1 Temporal Analysis

To study temporal aspects, we analyze the similarity and diversity of content requests over time.

Similarity. Let $\mathbb{D}_i$ be the set of content items requested during the day i. We then compute the Jaccard similarity $JS(i,j) = \frac{\|\mathbb{D}_i \cap \mathbb{D}_j\|}{\|\mathbb{D}_i \cup \mathbb{D}_j\|}$ for two given days i and j. Figure 6a shows the average **JS** of the content requested in a day with the following 28 days. Despite the overall decrease in similarity with time, there is a weekly pattern of high **JS** on the same day of the week.

To further investigate this diurnal pattern of user interests, we measure the **JS** of days in two consecutive weeks over four weeks. Figure 6b illustrates that there is clear difference in content requests received in weekdays and weekends. Moreover, Friday has the highest **JS** value with the following Monday rather than the weekend. This is most likely due to the cohort of commuters on weekends being different (typically non-returning) to the commuters during



(a) Similarity comparing to following days (b) Similarity week by week

Figure 6: Temporal Similarity of contents

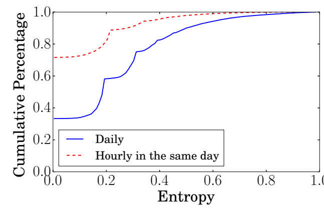


Figure 7: Daily and hourly entropy of the content

the weekdays (typically recurrent commuters). In addition, the closer the days, the higher the similarity, which is exemplified by the relatively higher **JS** values of the lower left triangle than top right, where the difference between the two days are less than 7 days.

Diversity. We define the entropy $E_T(i)$ of a content request i over different time slots (e.g. daily and hourly). To remove the bias towards highly popular time slots, the popularity of the request α_i^t during the time slot is considered instead of the absolute number of requests. Then,

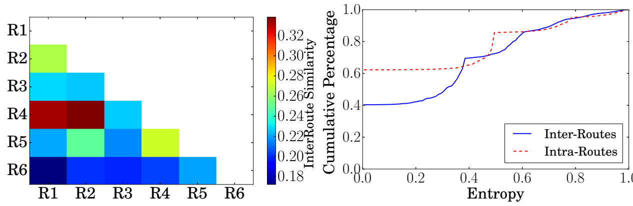
$$E_T(i) = -\sum_{t=1}^{n} \frac{(P_i^t \times \log P_i^t)}{\log n}, \text{where } P_i^t = \frac{\alpha_i^t}{\sum_{t=1}^{n} \alpha_i^t} \qquad (1)$$

P_i^t is the probability of request i coming from the time slot t. We calculate $E_T(i)$ for all i and Figure 7 shows the CDF of entropy for daily, for 35 days, and hourly, for a day. $E_T(i)$ of zero indicates that all requests for content i are received from one time slot, while $E_T(i)$ of one indicates that requests are generated from all time slots in equal proportions. Approximately 35% of the content are only requested in a given day and approximately 70% of them are only requested in the same hour. The results show that the temporal diversity of interests of commuters is significantly low and concentrated in smaller time slots.
Takeaway: *The daily similarity of popular content decreases in time while there is a periodic pattern of high similarity between content categories accessed on the same days of the week. Weekends also exhibit a very different workload/content set compared to weekdays.*

4.2 Spatial Analysis

We then study the spatial similarity and diversity, i.e. among routes and buses, while ignoring the time of requests for content.

Similarity. Similar to temporal Jaccard similarity measure, we calculate $JS(l,m)$ for routes/buses using the set of content items requested from a particular route/bus $l(\mathbb{R}_l)$. Figure 8a shows inter-route similarity during whole 5 week observation time period. Inter-route **JS** is illustrated in Figure 8a. R6 is very different to the other routes whilst R4-R1 and R4-R2 are the pairs with highest **JS** values. Notably, R6 is quite different from any other route. This uniqueness may



(a) Inter-route similarity (b) Entropy

Figure 8: Spatial similarity and diversity of content requests.

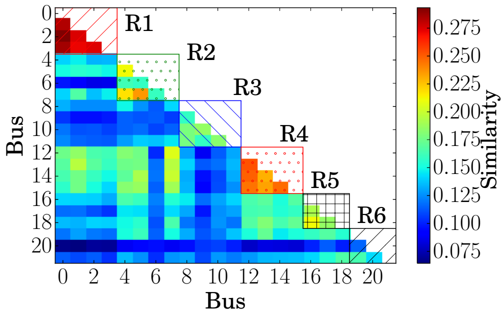


Figure 9: Intra-route and Inter-route similarity.

be due to the low proportion of recurrent commuters (Töcksfors city population is only 1000 residents) and longer travel time (120km route distance). Due to the shorter travel time, the commuters may opt to request the most popular content, e.g. most recent new item, and thus creating the high similarity among these routes.

Figure 9 depicts the intra-route similarity with **JS** values among all pairs of buses. More interestingly, the similarity among the buses in the same route are significantly higher than inter-route buses. For instance, all pairs of buses in R1 shows **JS** value more than 0.275. This result provides strong evidence to support the hypothesis that commuters traveling on the same buses/routes have similar interests.

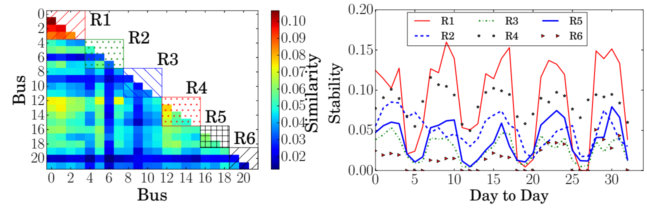
Diversity. We calculate the entropy $E_S(i)$ of a content request i among different routes and among the buses in the same route, again using Equation 1. The CDF of $E_S(i)$ in Figure 8b indicates that 40% of the content requests originated from a single route and approximately 60% of those requests generated from a single bus within the route. This surprising result further justify the strong spatial locality of user interests.

Takeaway: *Mobile commuters traveling in the same route, i.e. geographical area, have significantly similar interests compared to inter-route commuters.*

4.3 Spatio-Temporal Analysis

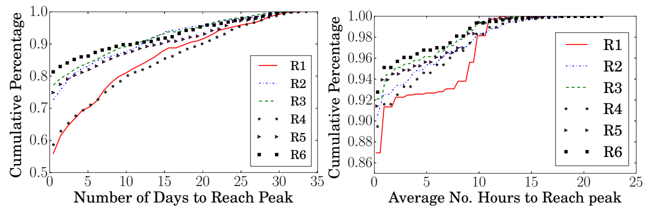
We now study the spatial and temporal correlation of content request patterns.

Similarity. Figure 10a shows the daily **JS** similarity across all 22 buses. Even though Figure 10a depicts similar pattern of **JS** distribution, the **JS** value drops during a day compared to Figure 9. For instance, similarity among R1 buses decreases from 28% (overall) to 11% (daily). This indicates that content may take more than one day to spread across a given route. To further analyze this behavior, we studied the stability of top most popular content and the time it requires for a content item to reach its maximum popularity.



(a) Similarity (b) Stability

Figure 10: Spatio-Temporal similarity and stability.



(a) Days to reach peak (b) Hours to reach peak

Figure 11: Time taken to reach peak popularity for content with more than 10 requests.

Figure 10b shows the portion of overlap $O(i,j)$ in the top most popular content between two consecutive days, where $O(i,j) = \frac{\|\mathbb{U}_i \cap \mathbb{U}_j\|}{\|\mathbb{U}_i\|}$ and U_i is the top most 10k popular contents during the day i. There is a periodic weekly pattern similar to request popularity, where R1 and R4 have considerably high stability than other routes. However, the highest stability value is about 15%, i.e. 15% of top most popular content will be in the same list in the next day. This again shows that popular contents life-span are short and majority of contents are popular temporally.

Overall, Figure 11 shows that 20-40% of content requires more than one day to reach its peak popularity. As observed earlier, R1 and R4 exhibits similar behavior with approximately 60% contents takes less than a day to reach its peak popularity while it is 80% for R6. Interestingly, approximately 90% of the content reaches its daily peak value within the first hour for all routes. Again, this result shows that commuter interests are highly concentrated in time.

Diversity. We consider active hours for each route/bus as those hours with at least 5% of the total requests of the day. As observed in Figure 4a, morning peak–7am, off-peak–12pm and afternoon peak–4pm are also considered to further compare the spatial and temporal diversity. The spatial entropy $E_{ST}(i)$ is calculated for a day and for peak/off-peak hours. Figure 12 shows that all E_{ST} values shifted towards zero compared to E_S values in Figure 8b. For instance, inter-route E_{ST} is less than 0.4 for approximately 92% of

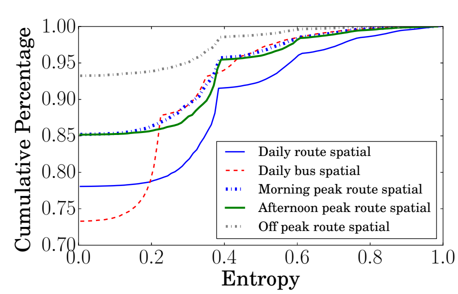
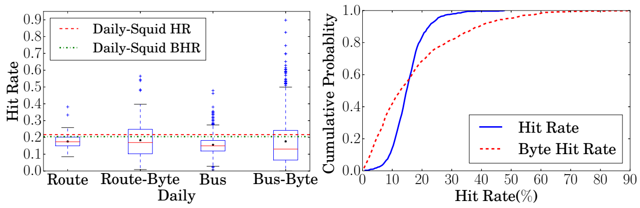


Figure 12: Spatio-Temporal diversity of content requests.

Table 5: Cache hit rate and byte hit rate of all time

Network Level	Overall Mean HR%	Overall Mean Byte HR%
Squid Server	31.80	27.47
All Routes	25.93	22.88
R1	30.60	26.98
R2	24.19	20.06
R3	24.17	23.10
R4	27.51	24.12
R5	24.74	25.20
R6	24.36	17.82
All Buses	21.45	20.18



(a) HR for routes and buses (b) Daily CDF for buses

Figure 13: Cache hit rate and byte hit rate distributions.

requests whereas it is 70% of requests for E_S. This indicates that when we narrow down the time frame, the commuter interests are further concentrated on to a smaller geographical area. In addition, spatial diversity is similar for morning peak and afternoon peak hours as shown in Figure 12. During both peak hours, 85% of requests are generated from only one route. However, during off-peak hours E_{ST} is less than 0.4 for approximately 98% of requests, i.e. commuters access unique unpopular content during off-peak hours.

Takeaway: *Interests of mobile commuters are highly concentrated in space and time.*

5. ON THE POTENTIAL OF CACHING

Since the device/user specific parameters matter in delivering exact content, the original request URL is used to uniquely identify a particular content[1]. Table 5 presents the cache hit rate (HR) and byte hit rate (BHR) at different network levels at the end of 5 weeks. Despite the fact that cache hit rates decreases as we move towards the edge of the network, some routes (R1 and R4) exhibit significantly higher hit rates due to the higher similarity and stability observed in Section 4.

We then further look into HR and BHR at the bus-level in a given day starting with an empty cache every day. Figure 13a compares the daily HRs of each of the bus routes. Due to the high similarity of interests within a bus, certain buses exhibits a very high performance improvement, as much as 47.9% HR and 89.8% BHR. More importantly, the 75^{th} percentile for both HR and BHR of the bus-level is larger than the route-level. Overall, the daily HR is between 10% to 20% for 80% of buses, while more than 40% of buses save more than 20% of daily bandwidth usage as shown in Figure 13b. Thus, the results show the high potential for performance improvement in caching at the edge of the network contrary to what has been reported in [6, 7], if we select user communities with similar interest such as commuters in public transportation systems.

[1]For this analysis, we assume unlimited cache storage and that every content item is cacheable, as the focus of the study is to investigate the impact of spatio-temporal correlation towards content caching.

6. CONCLUSIONS

In this paper, we studied the spatio-temporal correlation of content popularity among mobile commuters in public transportation systems by using a unique dataset from 22 inter-city buses for 5 weeks in Sweden. The analysis showed that, in contrast to previous studies, interests of mobile users are highly concentrated in spatially and temporally. We believe that these results can be effectively exploited to provide the desired quality of experience to truly mobile and challenged networks in public transportation systems. In this light, we analyzed the potential for content caching at the bus-level. Our results show a significant performance improvement due to the observed high spatio-temporal locality of interests.

7. REFERENCES

[1] Breslau, L., Cao, P., Fan, L., Phillips, G., and Shenker, S. Web caching and zipf-like distributions: Evidence and implications. In *Proceedings of INFOCOM'99.*, vol. 1, IEEE, pp. 126–134.

[2] Brodersen, A., Scellato, S., and Wattenhofer, M. Youtube around the world: geographic popularity of videos. In *Proceedings of WWW'12*, ACM, pp. 241–250.

[3] Cha, M., Kwak, H., Rodriguez, P., Ahn, Y., and Moon, S. Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on Networking (TON) 17*, 5 (2009), 1357–1370.

[4] Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019. Tech. rep., 2015.

[5] Dehghan, M., Seetharamz, A., He, T., Salonidis, T., Kurose, J., and Towsley, D. Optimal caching and routing in hybrid networks. In *Military Communications Conference (MILCOM), 2014 IEEE* (2014), pp. 1072–1078.

[6] Erman, J., Gerber, A., Hajiaghayi, M., Pei, D., Sen, S., and Spatscheck, O. To cache or not to cache: The 3G case. *IEEE Internet Computing 15*, 2 (2011), 27–34.

[7] Finamore, A., Mellia, M., Gilani, Z., Papagiannaki, K., Erramilli, V., and Grunenberger, Y. Is There a Case for Mobile Phone Content Pre-staging ? In *Proceedings of the 9th ACM conference on Emerging networking experiments and technologies* (2013), ACM, pp. 321–326.

[8] Jin, S., and Bestavros, A. Sources and characteristics of web temporal locality. In *Proceedings of Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 2000.*, IEEE, pp. 28–35.

[9] Kelleher, D. 95.6% of Commuters in the US put Company Data at Risk over Free Public Wi-Fi. Tech. rep., 2013.

[10] Li, Z., Lin, J., Akodjenou, M.-I., Xie, G., Kaafar, M. A., Jin, Y., and Peng, G. Watching videos from everywhere: a study of the pptv mobile vod system. In *Proceedings of IMC'12*, ACM, pp. 185–198.

[11] Li, Z., Xie, G., Lin, J., Jin, Y., Kaafar, D., and Salamatian, K. On the geographic patterns of a large-scale mobile video-on-demand system. In *IEEE Conference on Computer Communications (INFOCOM)* (Toronto, Canada, April 2014).

[12] McAfee. Customer URL Ticketing System, 2015.

[13] McNamara, L., Mascolo, C., and Capra, L. Media Sharing based on Colocation Prediction in Urban Transport. In *Proceedings of the 14th ACM international conference on Mobile computing and networking* (2008), pp. 58–69.

[14] Oxify. Oxyfi supplies on-board Internet to Norrtag, 2013.

[15] Press, A. Free Wi-Fi Catches on With NYC Subway Riders, 2012.

[16] Venkatramanan, S., Member, S., and Kumar, A. Co-Evolution of Content Spread and Popularity in Mobile Opportunistic Networks. *Mobile Computing, IEEE Transactions on 13*, 11 (2014), 2498–2509.

[17] Xu, Q., Huang, J., Wang, Z., Qian, F., Gerber, A., and Mao, Z. M. Cellular data network infrastructure characterization and implication on mobile content placement. *ACM SIGMETRICS Performance Evaluation Review 39*, 1 (2011), 277.