Supplementary File of the TPDS Manuscript:

Measurement and Analysis of an Internet Streaming Service to Mobile Devices

Yao Liu[1] Fei Li[1] Lei Guo[2] Bo Shen[3] Songqing Chen[1] Yingjie Lan[4]

[1]Dept. of Computer Science
George Mason University
{yliud, lifei, sqchen}@cs.gmu.edu

[2]Dept. of CSE
Ohio State University
lguo@cse.ohio-state.edu

[3]Vuclip
XinLab, Inc.
bshen@vuclip.com

[4]Guanghua School
Peking University
ylan@gsm.pku.edu.cn

Abstract—This supplementary file contents the supporting materials that were not included in the TPDS manuscript titled **Measurement and Analysis of an Internet Streaming Service to Mobile Devices** due to space limit.

1 Overview of Workload

With the exclusion of desktop/laptop traffic, Figure 1 gives an overview of the server side traffic in 30 days. Note that the left y-axis represents the total number of requests per day, while the right y-axis represents the total traffic volume per day. During this 1-month period, despite a small decrease in the middle, the number of the requests and the delivered traffic amount kept increasing, indicating the popularity of Vuclip.

Figure 2 shows the hourly mobile streaming access patterns in a day. The figure indicates that hourly accesses peak around 17:00 GMT. Furthermore, the total number of requests and the traffic volume served during peak hours almost double these in non-peak hours. Figure 3 further depicts the hourly pattern from Nov. 8th to Nov. 15th (a week). The figure shows clear peak and off-peak hourly patterns for each day. The figure shows some drop after Nov. 12th. It is likely due to the fact that Nov. 13th was a Saturday and Nov. 14th was a Sunday. We can observe the increase of accesses again on Monday.

Figure 4 shows the accesses during the 4-month period. In this figure, the x-axis represents the 120 days in our trace, the left y-axis represents the total traffic delivered from the server to the mobile devices over each hour, and the right y-axis represents the number of HTTP requests received every hour. Besides a clear daily pattern, we also find that during these 4 months, the number of requests received by the server has been increasing steadily, and so does the traffic delivered from the server. This indicates that Internet mobile video services are becoming more and more popular, and a study of the mobile video access pattern is in urgent need.

2 Characterization of Mobile Streaming Videos

2.1 Popularity of Mobile Videos

Figures 5(a) to (d) show the popularity distribution over different periods of time. Similar to Figure 9 in the TPDS manuscript, in these figures, the left y-axis is in powered scale while the right y-axis is in log scale. The x-axis is in log scale as well. As shown in the figures, the video popularity over different periods (two weeks, one month, two months, three months) all deviate from a straight line in log-log scale, meaning not a Zipf-like distribution. Instead, they can roughly be fitted with a stretched exponential (SE) distribution. From Figures 5(a) to (d), we find that parameter a increases from 0.091 (two weeks) to 0.106 (three months). This confirms that the parameter a in an SE distribution increases over time [1]. On the other hand, the stretch factor c remains an invariant, as the file size distribution of the workload also remains stable during this time period.

2.2 Popularity of Different Video Versions

We have shown in Section 4.2 of the TPDS manuscript that the daily video popularity follows a Zipf-like distribution. However, as discussed before, each video may be accessed by very diverse mobile devices, resulting in multiple transcoded versions. We thus further examine the popularity distribution of all versions accessed on Nov. 1st where each version is counted as a distinct

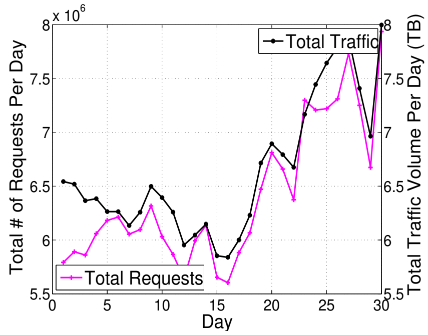


Fig. 1. Daily Accesses in Nov. 2010

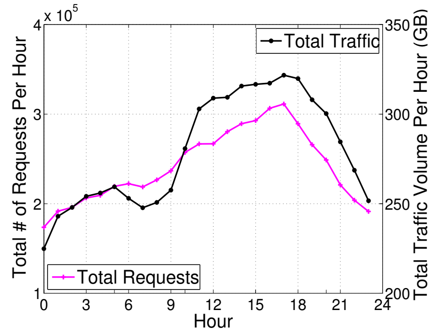


Fig. 2. Hourly Accesses On Nov. 1st 2010

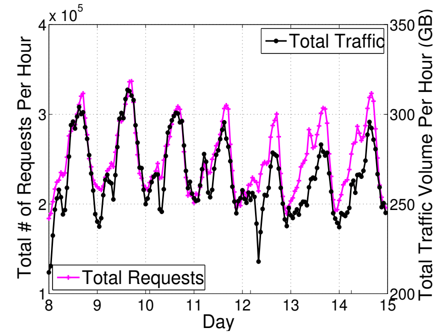


Fig. 3. Weekly Accesses from Nov. 8 to Nov. 15 2010

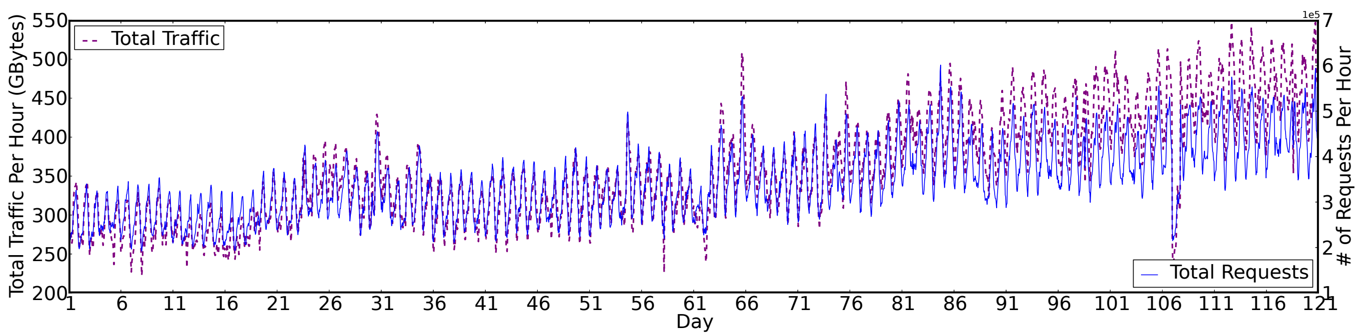
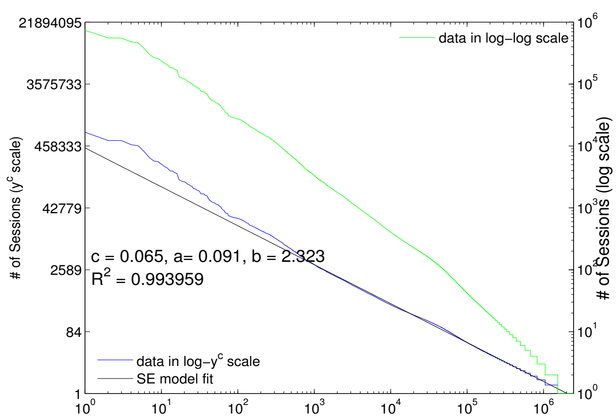
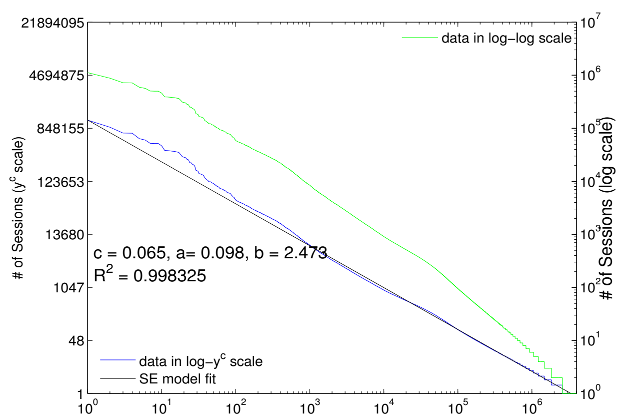


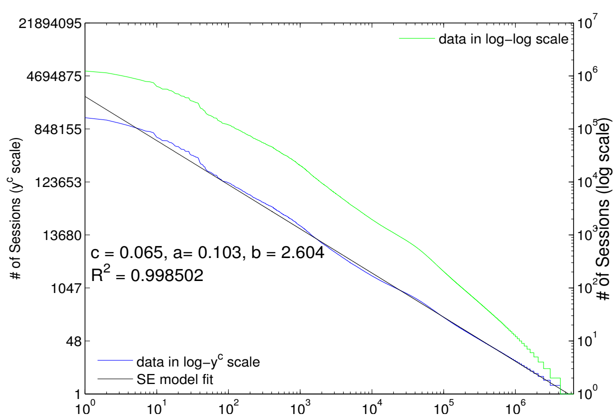
Fig. 4. Hourly Accesses Pattern over 4 Months



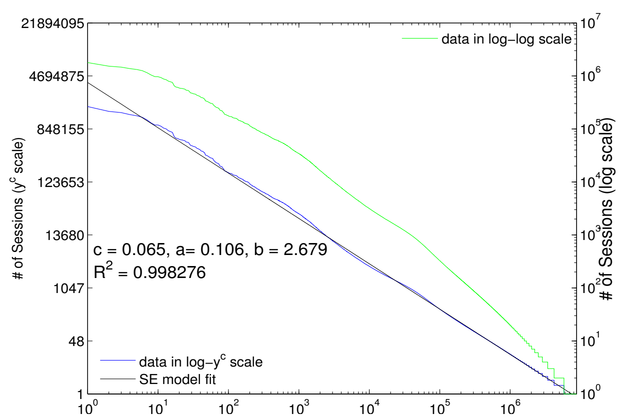
(a) Two Weeks (Nov. 1 to Nov. 14)



(b) One Month (Nov.)

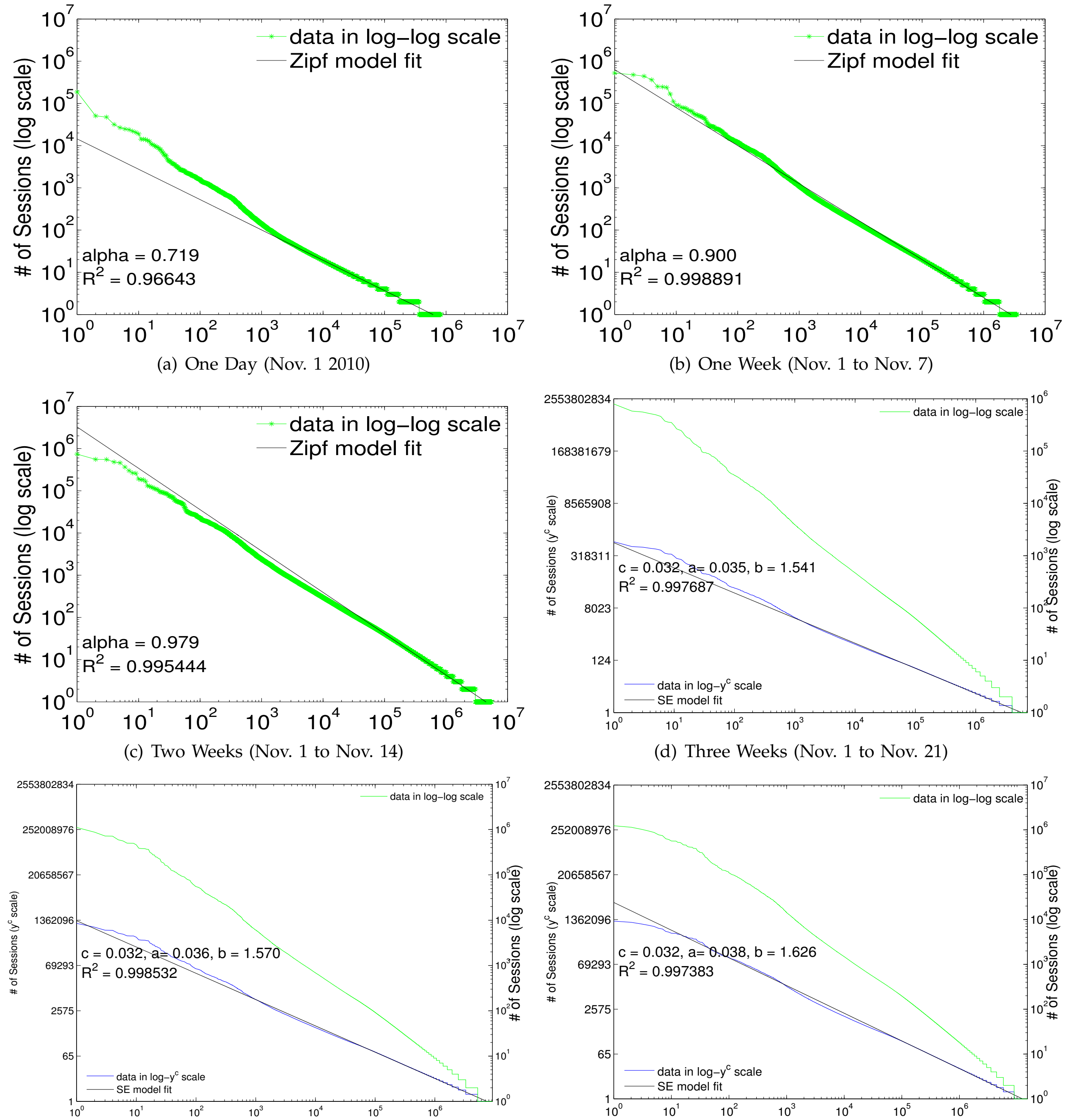


(c) Two Months (Nov. to Dec.)



(d) Three Months (Nov. to Jan.)

Fig. 5. Video Popularity Distribution



(a) One Day (Nov. 1 2010)

(b) One Week (Nov. 1 to Nov. 7)

(c) Two Weeks (Nov. 1 to Nov. 14)

(d) Three Weeks (Nov. 1 to Nov. 21)

(e) One Month (Nov.)

(f) Two Months (Nov. to Dec.)

Fig. 6. Version Popularity Distribution

object. Figure 6(a) shows that when different versions are considered as different objects, the popularity cannot be well-fitted with the Zipf distribution. On the one hand, due to the increased number of video versions (2.31 versions per video on average over 1 month), the skewness factor α decreases from around 0.95 to 0.7. On the other hand, as shown in the figure, the accesses to the Top-1000 versions are much larger than what Zipf predicts, indicating significant deviation from Zipf-law.

Over medium term (e.g., one week and two weeks), we find that the popularity distribution can be well fitted with the Zipf distribution, with the goodness of fit R^2 very close 1 as shown in Figures 6(b) and (c). But in longer terms, similar to the popularity of different videos, the popularity pattern of different versions converges into SE distribution as shown in Figures 6(d), (e) and (f). This is because the popularity distribution is non-stationary over long terms, and the most popular versions cannot keep up with the same popularity as new videos and new versions join the system. For detailed mathematical analysis, please refer to Section 4.2 in [1].

TABLE 1
Summary of 1-month Video Accesses

Average Daily New Videos	92 K
Average Daily Accesses Videos	502 K
Percentage of Daily New Videos	18%
Average Daily Accessed Videos Difference	292 K
Percentage of Average Daily Difference	58%
Average Daily Requested Sessions	3512 K
Total Accessed Videos in 30 Days	4052 K
Percentage of New Videos in 30 Days	68%

2.3 Popularity Evolution

We have shown that mobile users' daily accesses for mobile videos are highly concentrated, but monthly accesses patterns are flatter. In this subsection, we further examine how video popularity changes over time, aiming to shed light on such popularity changes. We first study the commons between accesses in consecutive days, and then, more specifically, we consider the temporal locality characteristics of these accesses.

Table 1 summarizes the daily accesses and the corresponding videos that were requested in the system based on our 1-month log. According to our analysis, 292K out of 502K (58%) unique video clips accessed daily are not accessed in the previous day on average. Among these 292K video clips, about 92K are new video clips. This indicates that about 18% video clips accessed every day are new. The rest 200K (40%) are unpopular ones that were in the system, but were not frequently accessed. In total, new video clips account for about 68% of total unique video clips accessed during 30 days. Since new video clips are generated at a high rate of 18%, this confirms the implication of SE-distribution that a monthly static caching scheme may not be so efficient as a more frequently updated one.

Unlike traditional video on-demand streaming systems, Vuclip has a larger repository as well as a faster new content generation rate. We next examine if temporal locality is helpful in predicting what will be popular in the future in such a highly dynamic system.

Figure 7 keeps track of the Top-100, Top-500, Top-1000, Top-2000, Top-5000, and Top-10000 videos that were accessed the most on the first day of our trace, Nov. 1st, 2010. We examine how the popularity of these top videos would evolve over these 4 months. We find that the Top-100 video list has the fastest drop-out rate, as only fewer than 20% of the videos would remain in the list after 13 days, while it takes 54 days for Top-500, 80 days for Top-1000, and 109 days for Top-2000 video lists. After 119 days, about 30% of videos on Top-5000 lists and 32% of videos Top-1000 list would still remain on such lists.

Compared to the change of popular videos, Figure 8 shows the changes of popular versions. We find that the drop-out rate for Top-100 and Top-500 version lists are similar to Figure 7. However, the Top-1000, Top-2000, Top-5000, and Top-10000 version lists are seeing faster drop-out rate than that in Figure 7. As a result, only fewer than 20% videos on these lists would remain. This indicates version popularity changes more dynamically than video popularity.

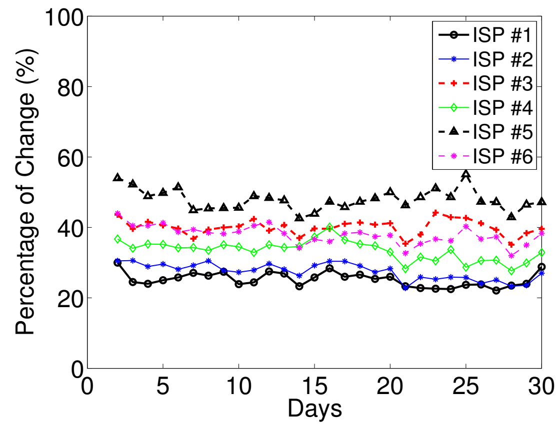


Fig. 11. % of Change in Top-1000 Videos

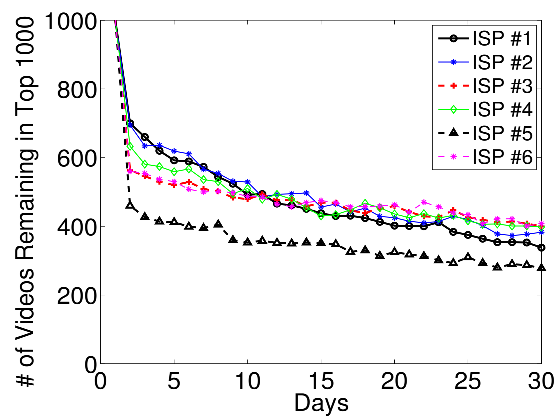


Fig. 12. # of Videos Remaining in Top-1000

Having shown how videos' popularity would change over time, we further study how the list of popular videos changes everyday. Figure 9 shows the percentage of videos in the top lists that are different from the previous day. We find that such a change rate varies between 15% and 37%, which is faster than the traditional VoD system [2]. In comparison, Figure 10 shows that the change rates of Top-100 and Top-500 version lists are similar to the top video lists, while the change rates in Top-1000, Top-2000, Top-5000, Top-10000 version lists are higher than the corresponding top video lists.

3 Characterization of Accesses based on ISPs

3.1 Popularity Evolution within ISP

Figure 11 shows the percentage of change in Top-1000 requested videos in the 6 ISPs. Compared to Figure 9, it can be seen that the shifting of user interest within ISPs (varying between 25% and 50%) is faster than site-wide

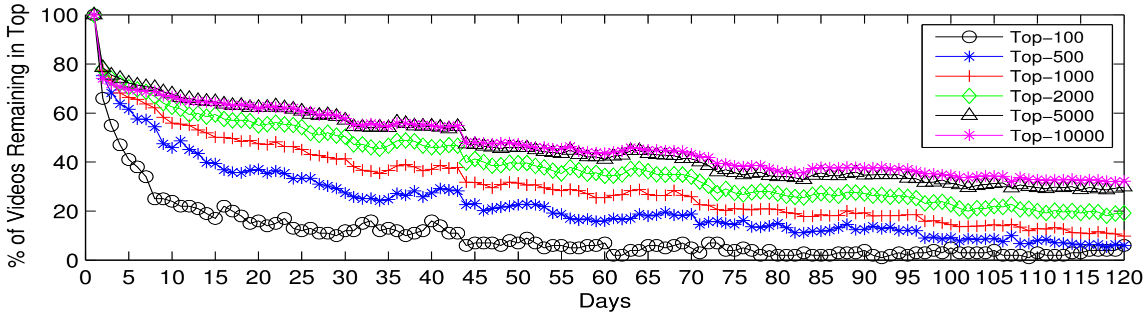


Fig. 7. % of Videos Remaining in Top

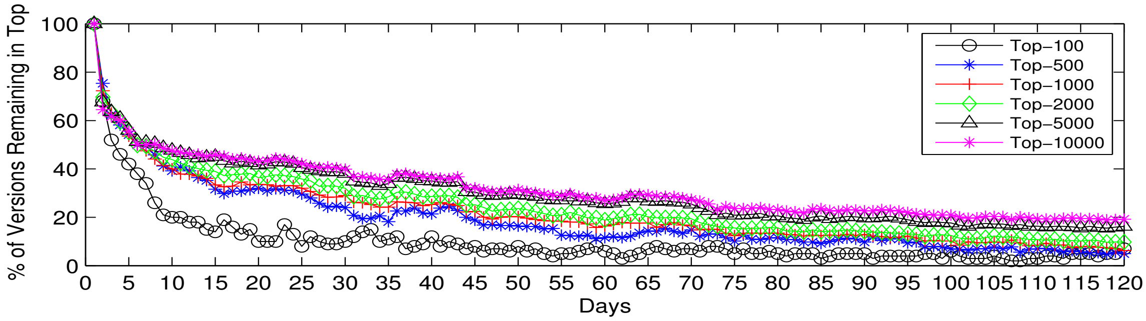


Fig. 8. % of Versions Remaining in Top

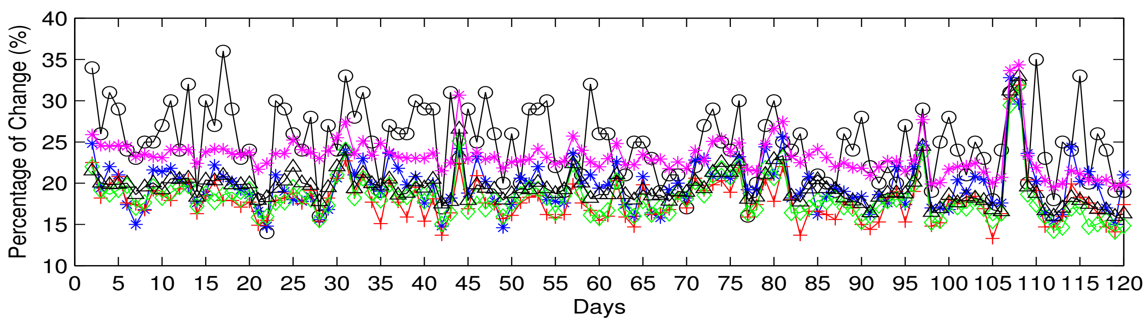


Fig. 9. % of Change in Top Video Lists

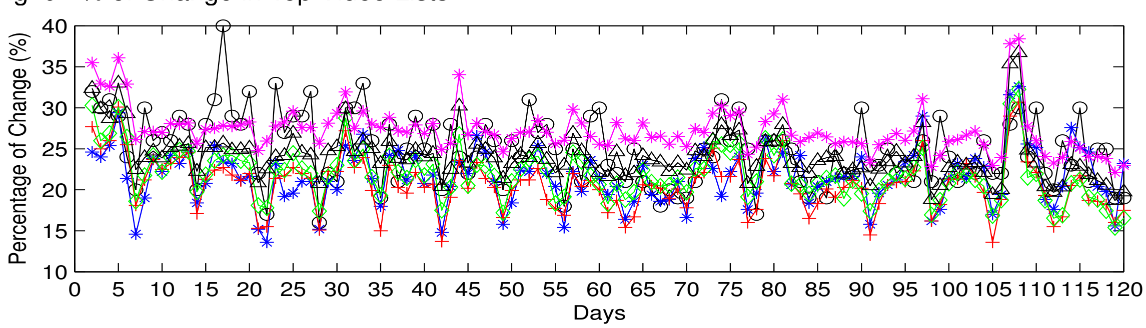


Fig. 10. % of Change in Top Version Lists

(around 20%). This indicates if prefetching is employed at the ISP side, it would be less effective in some ISPs. We further examine the object hit ratio if the Top-1000 accessed videos are prefetched into the ISP's proxy. Table 2 shows the results based on our 1-month trace.

It is surprising that the effectiveness of prefetching can be largely different across the six ISPs. For example, the hit ratio is higher than 0.7 for ISP #1 and ISP #2, and could be as low as around 0.2 for ISP #3, #5, and #6. The underlying reason, as we presume, is two-fold. On one hand, the percentage of change is relatively small for ISP #1 and #2 (less than 30%), so that more videos would remain in top accessed list. On the other hand, as we have shown in Figure 10 of the TPDS manuscript,

TABLE 2
Prefetching Top-1000 Videos the Previous Day

ISP	Hit Ratio
#1	0.71
#2	0.75
#3	0.23
#4	0.33
#5	0.22
#6	0.21

the daily access pattern of ISP #1 and #2 are much more concentrated than the other four.

However, despite the higher rate of change in top videos in ISP #3, #4, #5, and #6, the long-term temporal locality across all 6 ISPs as shown in Figure 12 are similar to that site-wide (Figure 7): around 30% to 40% videos would remain in Top-1000 list throughout the 29 days. And the number of videos remain in Top-1000 in ISP #3, #4, #5, and #6 becomes even more stable after the initial days of change, even though they have higher initial purge out rate.

It remains unclear why user interests in the four ISPs changes faster. Our conjecture is that this is related to the different presentation models and different recommendation systems different ISPs use to present the Vuclip site to the users. We will leave this for future work.

References

[1] L. Guo, E. Tan, S. Chen, Z. Xiao, and X. Zhang, "The Stretched Exponential Distribution of Internet Media Access Patterns," in *Proc. of PODC*, 2008.
[2] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng, "Understanding User Behavior in Large-Scale Video-on-Demand Systems," in *Proc. of EuroSys*, 2006.