

Watching Television Over an IP Network

Meeyoung Cha* Pablo Rodriguez[†] Jon Crowcroft[†] Sue Moon[‡] Xavier Amatriain[¶]
*MPI-SWS [¶]Telefonica Research [†]University of Cambridge [‡]KAIST

ABSTRACT

For half a century, television has been a dominant and pervasive mass media, driving many technological advances. Despite its widespread usage and importance to emerging applications, the ingrained TV viewing habits are not completely understood. This was primarily due to the difficulty of instrumenting monitoring devices at individual homes at a large scale. The recent boom of Internet TV (IPTV) has enabled us to monitor the user behavior and network usage of an entire network. Such analysis can provide a clearer picture of how people watch TV and how the underlying networks and systems can better adapt to future challenges. In this paper, we present the first analysis of IPTV workloads based on network traces from one of the world's largest IPTV systems. Our dataset captures the channel change activities of 250,000 households over a six month period. We characterize the properties of viewing sessions, channel popularity dynamics, geographical locality, and channel switching behaviors. We discuss implications of our findings on networks and systems, including the support needed for fast channel changes. Our data analysis of an operational IPTV system has important implications on not only existing and future IPTV systems, but also the design of the open Internet TV distribution systems such as Joost and BBC's iPlayer that distribute television on the wider Internet.

Categories and Subject Descriptors

C.2.3 [Computer Systems Organization]: Computer-Communication Networks—*Network Operations*; C.4 [Computer Systems Organization]: Performance of Systems—*Measurement techniques*

General Terms

Design, Measurement

Keywords

IPTV, channel switching behavior, network measurement

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IMC'08, October 20–22, 2008, Vouliagmeni, Greece.

Copyright 2008 ACM 978-1-60558-334-1/08/10 ...\$5.00.

1. INTRODUCTION

Since the 1950's, television has been a dominant and pervasive mass media; it is watched across all age groups and by almost all countries in the world. Over the years, television has transformed itself into a new media. The number of channels has increased from a few free-to-air broadcasts to several hundreds for cable, satellite, and Internet TV networks, that transmit more channels to each user. The video signal itself has changed from black & white and color analog to high-quality digital stream.

Many technological advances were produced by trying to meet user needs and expectations in such a widespread media. For example, the large number of users that concurrently watch TV sparked the use of IP multicast by network operators to provide Internet TV (IPTV) services with low transmission costs. And now traditional media is converging with newer Internet-based services (e.g., Joost [1], Zattoo [2], Livestation [3], and BBC's iPlayer [4]). In such a context, IPTV is a promising starting point for research because it opens up the door for many innovations while still keeping its roots in the traditional TV watching paradigm.

Despite the widespread usage of television and its importance to emerging applications, the ingrained TV viewing habits are not completely understood. Nielsen Media Research [5] spearheads a long-standing research effort to estimate TV viewing behaviors through monitoring and surveys. However, due to the difficulty of instrumenting monitoring devices at individual homes, it is hard to monitor user behavior across the entire network. Sophisticated methods such as stratified sampling, systematic sampling, and cluster sampling are used to find a representative set of users and extrapolate their behaviors to the entire population [6].

With the recent large-scale deployment of commercial IPTV systems, the IPTV workloads add a new dimension towards understanding TV viewing habits. The bidirectional communication in IPTV systems gives more visibility on TV viewing activities across an entire network. At the same time, the large user base provides a clearer picture of how people watch TV across different groups of users (e.g., location, content genre, etc).

In this paper, we present the first in-depth analysis of IPTV workloads using anonymized traces obtained from a live IPTV network. Our trace spans a six-month period and includes network activity information generated by roughly a quarter million households¹, an order magnitude larger than that of Nielsen. We characterize several properties

¹We are not able to differentiate individuals in the same household sharing a TV.

of aggregate viewing sessions, channel popularity dynamics, and content locality, which help content producers and service providers create more appealing programs and a better broadcasting schedule. We also identify the browsing patterns of TV users to find content and characterize the user arrival and departure patterns across channels, which could be used as an input to design future IPTV systems. Our analysis focuses on aggregate user behavior to understand the impact in the network, but does not characterize viewing patterns of individual users.

The highlights of our findings could be summarized as follows:

- 1) Over 60% of channel switching events happen within 10 seconds, indicating user focus is very volatile.
- 2) Channel popularity changes widely over a day and popularity of top channels is reinforced during peak times. Popularity distribution is Zipf-like for popular channels and it decays fast for non-popular ones.
- 3) When watching TV, attention span is content genre-dependent and ranges from 5 to 21 minutes. Daily viewing patterns are also heavily dependent on genre.
- 4) We observe strong differences in the viewing patterns of different regions within a country. Across thousands of local communities, the viewing patterns vary up to 20%.
- 5) Channel switching rates depend on program popularity; people change channels less when watching popular channels. Channel switching events are predominant (over 60%) and can have a major impact in user's experience as well as network performance.
- 6) The set of channels people watch (for longer than 1 minute) follows non-sequential channel numbers, while 62% of channel changes are sequential.

Our findings, based on non-sampled data, have several important implications for the design of both future television and the open Internet TV distribution systems. In particular, users' channel selection process can have a major impact on the network, as significant channel changes generate unnecessary network load. The design of a system that supports fast channel switching and helps users find the right channel quickly is imperative to both improving end user experience and minimizing the impact in the network. There are also observe high spikes in departure patterns, which could affect the design of peer-to-peer (P2P) IPTV systems. Finally, the users' tastes are not as homogeneous as one would expect over time, geography, and content, and therefore, our findings can provide important input towards designing efficient IPTV channel recommendation systems.

The rest of the paper is organized as follows. We describe the context and background of IPTV in §2. We introduce the dataset and our assumptions in §3. §4 presents our characterization of the viewing session, channel popularity, and geographical locality. §5 focuses on the dynamics of channel switching behaviors. In §6, we discuss the implications of our findings. We present related work in §7, and discuss issues for further work and conclude in §8.

2. CONTEXT AND BACKGROUND

We introduce the IPTV services and its architecture, and summarize the previous studies on TV habits.

2.1 IPTV market

The IPTV market has grown rapidly over the past few years. As a whole, an estimated 13.2 million households received IPTV services in 2007 worldwide and this number is projected to grow 52.2% annually through 2012 [7]. Major IPTV players are the telcos (telecommunication service providers) that run national backbones. There are other types of IPTV services, that are not based on high-speed real-time streaming but primarily on downloading movies and TV programs before viewing (e.g., Apple TV [8], Vudu [9], Amazon's Unbox for Tivo [10]). Despite the huge success, no large-scale studies have been published on the viewing habits from IPTV services. In this paper, we study the intrinsic characteristics of what and how people watch TV using traces collected from one IPTV live operation.

2.2 IPTV service architecture

Most telco's IPTV architecture is based on "walled-gardened," well-provisioned IP multicast infrastructure [11, 12]. Figure 1 illustrates a typical service architecture, where customers subscribe to IPTV, IP phone, and Internet access. A network device called Digital Subscriber Line Access Multiplexer (DSLAM) aggregates traffic from hundreds to thousands of users and connects to the high-speed IP backbone. For IPTV distribution, the TV head-end streams all the live broadcast TV channels towards DSLAMs through bandwidth-provisioned multicast trees in the backbone.

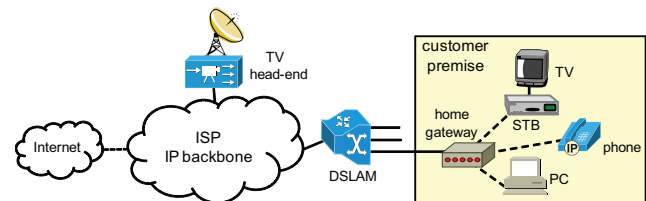


Figure 1: Telco's IPTV service architecture

Due to limited access bandwidth at the last mile (i.e., from a DSLAM to the home gateway at the customer's premise), not all TV channels are delivered to customers simultaneously. The IPTV network studied in this paper streams one channel to each customer. However, multiple channels can be delivered to support multiple TVs at home. The IPTV set-top-box (STB) translates a channel switching event from a user into a pair of Internet Group Management Protocol (IGMP) messages: one to alert the user's departure from the multicast group of a channel and the other to join another multicast group of a new channel. Upon receiving IGMP messages, DSLAMs extend or prune multicast trees dynamically towards the customers' STBs.

2.3 Studies on TV viewing

The Nielsen Media Research provides a wide set of statistics on TV viewing and program ratings based on a sample of population [5]. The Nielsen technicians install metering equipment on TV sets at sampled home, and record minute-by-minute viewing of users. These records are collected on a daily basis and mapped with the broadcast program schedules to produce the Nielsen TV ratings. Their sampling devices now gather traces from approximately 25,000 people

from 10,000 homes in the United States. One of their metering equipments called “peplemeters” [13] assign a personal viewing button for each family member in a sample household, which is matched to that person’s age and sex. This is to identify viewing habits of specific groups (e.g., young men in their 20s and 30s).

However, there are known issues with sampling methods such as the ones used by Nielsen. Selected households are not accurate in terms of demographic and regional weighing and since viewers are aware of being monitored, it can lead to bias in viewing habits. Furthermore, regardless of the increasing incentives offered by audience analysis companies, over 30% of the selected households refuse to be part of the studies [14]. There is no available information of how this may bias the statistics. In contrast, IPTV providers can monitor behaviors of all the users closely and continuously without having to install extra monitoring devices. However, IPTV providers lack information about what the individual members of each family are watching. Our study is also limited by the absence of such information.

3. DATASET AND ASSUMPTIONS

We describe our methodology to collect and process logs and discuss our assumptions. Throughout this paper, we refer to individual households as **users** and use the term **switching** to refer to the user action of changing channels.

3.1 Trace collection

We obtained a collection of IPTV channel switching logs from an operational backbone provider. The trace spans six months from May to October, 2007 and records the IGMP messages on the channel changes of 250,000 users. The log includes the following:

- **Timestamp** in units of seconds
- IP address of the DSLAM
- IP address of the **set-top-box (STB)**
- IP address of the **multicast group (channel)**
- **Multicast option** of join or leave.

Some multicast groups are used to manage set-top-boxes (e.g., bootstrapping and software upgrades). We pre-process the log by excluding events of non-video multicast groups. We then chronologically sort IGMP join messages. This allows us to identify which program a user was watching, for how long, and how one switched channels. Due to the scale of our trace, we present results that are averaged across the trace period or from a representative day.

There are 150 channels, which are grouped by the IPTV provider as follows: First, there are national, free-to-air broadcasted channels, which we refer to as **free**, and **local** channels produced by the regional states. Then there are **audio**, **satellite**, and pay-per-view (**ppv**) channels. The rest are cable TV-like content including children’s programs (**kids**), documentaries (**docu**), movies (**cine**), **sports**, **news**, and **music** channels. We refer to the collection of comedy, soaps, and reality shows of the remaining channels (e.g., Fox, Paramount comedy) as **mixed**.

Our data does not capture other performance related metrics such as the network latency, jitter, and loss of the

IPTV streams. Also, at the time of data collection, the deployed IPTV system did not support digital video recorders (DVRs), either in the network or in the set-top-box, that allow people to store favorite shows and view later. We leave the studies of IP multicast throughput and behavioral changes due to DVRs as future work.

3.2 Assumptions on user modes

Before we proceed our analysis and in order to better understand the user behavior, we will divide user activity into three different modes, which we will call **surfing**, **viewing**, and **away**. Our away mode indicates whether the user has turned off the TV or the TV is left on without anyone watching it. Our trace does not include information to differentiate these two options, because most users leave the IP set-top-box on and continue receiving multicast stream, even when the television is off.

To define these three modes, we examine the channel holding times, i.e., the time interval between channel switchings. Figure 2 shows the cumulative distribution of channel holding times, where the slope of the curve changes around 10 seconds and again around 1 hour, indicating natural divisions of the times between these three different modes. Although these divisions may somehow look arbitrary, they coincide very well with those divisions made by the traditional media research. For example, the Nielsen Media Research demarcates viewing events by the minute [5]. MSK Research in Korea uses 15 seconds for viewing period; TNS Media in Korea considers the channel that a user spent most time during every 1 minute as the viewed channel.

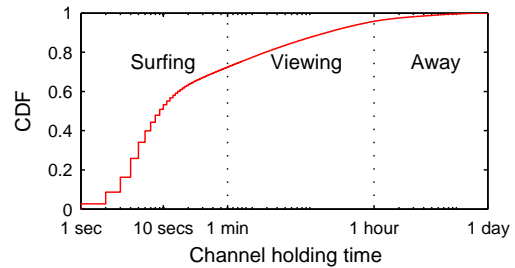


Figure 2: Inferring three user modes; 60% of channel holding times are shorter than 10 seconds

We set the viewing threshold (or viewing period) to 1 minute to be consistent with the Nielsen Research and to be able to compare the data better. We set the away threshold (or away period) to 1 hour. The most popular channels in our IPTV system are free-to-air, which are paid by advertisements, and therefore it is hard to find any single program longer than 1 hour with no interruptions. Some users may watch TV for more than 1 hour without channel changes, especially when watching movie channels or when checking the Internet at the same time as watching a program. As we do not have information about the detailed user activity, we focus on the “active online viewers”—the users who denote some activity during a 1 hour period.²

²Although we borrow the concept of “active viewer” from traditional media research [15], we do not mean to imply that our threshold-based method is a good measure of viewing activity.

To validate our findings, we have repeated most of our analyses with different thresholds. The results were not sensitive to changes in the thresholds, for instance, from 1 minute to 5 minutes or from 1 hour to 2 hours. Due to space limitations, we do not present those results.

4. HOW, WHEN, AND WHAT PEOPLE WATCH

In this section, we characterize several properties of TV viewing behaviors, in particular:

- *User session characteristics*: when people watch TV and what is their attention span across genre.
- *Channel popularity and dynamics*: how user interests are spread across channels over time.
- *Geographical locality*: whether users in the same region or DSLAM show similar viewing patterns.

4.1 High-level viewing characteristics

We present our findings on viewing characteristics.

4.1.1 Number of simultaneous online users

We examine the number of concurrent online users based on our notion of users on- and off-line. Figure 3 shows the day-of-week trend over a representative trace period. The start of each day is marked in the horizontal axis. The number of viewers in Figure 3 shows some interesting patterns. First, the graph shows a diurnal pattern with two large peaks around 3PM and 10PM, and a small peak around 8AM. They correspond to dining hours of the population. Other video systems showed similar diurnal patterns with two peaks [16,17]. We will shortly examine how these on-line users are distributed across individual channels.

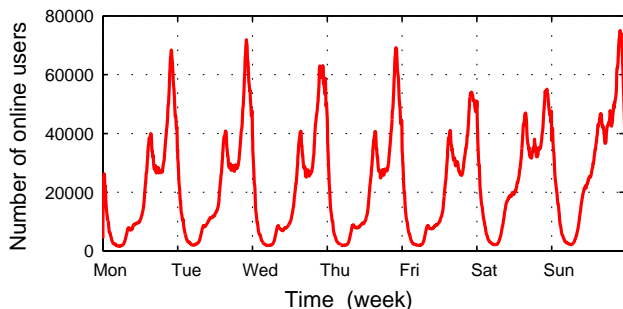


Figure 3: Number of viewers over a week period

Unlike patterns of Internet usage [18], Friday and Saturday consistently had the lowest evening peaks within the week. On weekends, the number of viewers ramps up from early morning, which was due to the increase in viewing of children’s programs. Further details reveal that on weekends the number of distinct viewers increase by 5% and the total time spent on TV increases by 30%. Such a trend has been also reported in a survey of the UK workers about their leisure time use [19].

4.1.2 Session characteristics

An average household watches 2.54 hours and 6.3 distinct channels of TV a day. The average length of time users watch TV each online session is 1.2 hours, and each on-line session consists of multiple surfing and viewing events. According to the traditional survey results reported by Nielsen [20], an average household watches television for approximately 8 hours and 14 minutes per day and an individual watched 4 hours and 34 minutes a day. Other surveys report that an average household in Australia watches 3 hours of TV a day [21] and an average person in UK watches TV for 3 hours and 38 minutes a day [22].

Across a day, the median and the mean channel holding times are 8 seconds and 14.8 minutes, indicating that a significant fraction of user activities are short-intervalled (i.e., surfing) while some last much longer (i.e., away) as we saw in Figure 2. The median and the mean channel holding times are 6 and 10 seconds when surfing, 4.8 minutes and 10.7 minutes for viewing activities, and 3.1 hours and 5.7 hours when away, respectively.

Investigating the session durations more closely, we examine the times spent on watching television before changing a channel (i.e., sojourn times). Using a representative day trace, we show the frequency of sojourn times of all online users in Figure 4, in a log-log scale. The frequency of a TV watching duration increases from 1 to 4 seconds, then drops very quickly. Most channel changes happen 4 seconds after the previous channel change action. The increase in the frequency from 1 to 4 seconds may be affected by multiple issues. For example, the effective channel switching time (or so-called the zapping time) may vary per household depending on the distance of the set-top-box to the local DSLAM. Also, individuals vary in the ability to quickly recognize a program. However, the exact impact of these issues needs further investigation.

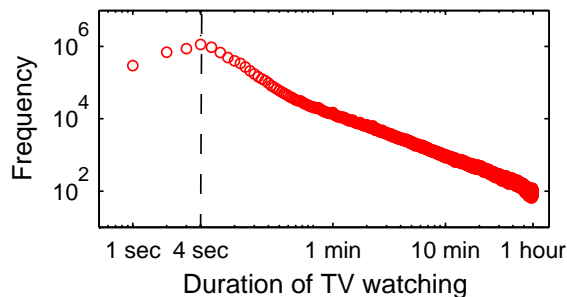


Figure 4: Duration of time watching each channel

The graph after the 4-second mark follows a power-law distribution [23,24]. This reflects an interesting tendency of users to continue watching the same channel: the probability of continuing to watch a channel increases proportionally with the length of time already spent watching that channel. This is in contrast to the exponential distribution, which has generally represented human activity in environments, such as wireless access [25] and telephone call [26] durations, and reflects the memoryless property that future activity is independent of past activity. According to Vazquez [27], the power-law pattern in humans could be due to people having some perception of their past activity rate (i.e., not memory-

Table 2: Breakdown of popularity across genre (probability of a viewer watching each genre)

genre	free	mixed	kids	docu	local	cine	sports	music	news	audio	the rest	total
viewing prob.	38.6%	21.5%	12.5%	6.6%	4.9%	3.9%	3.8%	2.3%	1.0%	0.3%	4.6%	100%
num. channels	6	19	7	12	17	6	8	11	13	15	36	150

*Genre categorized “the rest” includes pay-per-view, satellite, and promotional channels.

Table 1: Time spent on surfing and viewing a single TV program grouped by channel genre

perc.	news	kids	music sports	free mixed	docu cine
50th	6 sec	6 sec	7 sec	9 sec	9-11 sec
90th	3 min	10 min	2-3 min	6-7 min	5-6 min
95th	9 min	21 min	5-8 min	17-18 min	15-17 min

less) and thereby reacting by accelerating or reducing their activity intensity [28, 29].

4.1.3 Attention span

When watching television, people browse through a set of on-air TV programs until they find something interesting. The behavior of channel selection involves the following two steps: (a) browsing content to decide whether to continue or stop streaming and (b) switching through multiple channels for repeated browsing, until a desired channel is found. In Figure 2, we observed that over 60% of user action is related to channel surfing. To understand in detail how long a single sample takes, we summarize the characteristic channel holding times for a few channel genres in Table 1.³

The 50th percentile values range from 6 to 11 seconds, indicating that viewers watched only for a very short while and changed the channel. The median duration is slightly longer for **free**, **mixed**, **docu**, and **cine**. The 90 and 95th percentile values reflect the viewing durations and show larger variations ranging from 3 to 21 minutes. At the 90th and above percentiles, the viewing durations of **music** and **sports** are much shorter than those of **kids**, **free**, **mixed**, and **cine**—or conversely, the attention span is much shorter when watching **music** or **sports**.

4.1.4 Time spent on each genre

We now investigate the fraction of time user spent on these genres. Table 2 summarizes the percentage of time spent on each genre, over the trace period. **free** and **mixed** genres account for over 50% of TV viewing instances. [30] reports that these genres (identified as news and public affairs, comedy, and drama in the report) similarly account for over 60% of TV viewing in Canada. Regional content is also popular, considering that each region has one or two **local** channels. **Kids** and **docu** programs are also popular. There are significant differences in our findings of channel genre popularity and those reported by sampling statistics for the US population [31]. For instance, [31] reports **kids** and **docu** channels to almost double the percentage in viewing share we find in our analysis. While this could well be a trend in different social behaviors in the analyzed populations, it could also be

³We do not have demographic information about the populations watching individual channels or how different populations exhibit different social behavior.

attributed to a change in viewing behavior when households are under the scrutiny of a tracking device.

4.2 Channel popularity and dynamics

The current IPTV system continuously sends out multicast streams of all channels throughout the destinations in the backbone (i.e., DSLAMs). However, such fixed distribution may no longer be possible if and when the number of channels increases. In certain cases, the backbone needs more bandwidth to deliver channels in multicast rather than in unicast. One strategy to minimize the network cost in the backbone is to use multicast for some channels, while carrying the rest as unicast (or P2P distributed fashion [32]). Accurate measurement of channel popularity is critical in designing such systems. In this section, we study channel popularity from three different angles. First we analyze its distribution to understand how user interests are spread across channels. We then investigate how fast channel popularity changes over time, and what causes such dynamics.

4.2.1 Popularity distribution

We start with channel popularity distribution. Figure 5(a) shows the aggregate fraction of viewers accounting to the least x -th popular channels. The horizontal axis represents the ranking of channels normalized between 0 and 100, from the most popular to the least popular. The top 10% of channels account for nearly 80% of viewers, which reflects the Pareto principal or 80-20 rule. This is consistent across different times of the day. It is important to note that our notion of channel popularity is from the system’s perspective, not from the individual user’s. The audience base changes over the course of a day, i.e., different households watch TV at different times of the day. While the audience demographic changes, we consistently find that the Pareto principal still holds throughout the day as shown in Figure 5(a).

Figure 5(b) shows the popularity distribution as a function of the channel ranks against the number of viewers at a short time scale. The channel rank of 1 indicates the most popular channel and the unpopular channels are put at the tail. We calculate the effective number of viewers by the fraction of time a user spent on each channel over a minute period (e.g., a user watching a channel for 30 seconds is counted as 0.5 viewer). We take the average effective number of viewers per channel over 15 consecutive minutes. The plot shows a straight line waist for top ranked channels (below 30) and a truncated tail. This indicates that popularity is Zipf-like for top channels and it decays fast for non-popular ones. Other live video streaming systems have showed similar trends [16, 33, 34].

In our IPTV system, some channels also fail to attract enough viewers compared to a normal Zipf distribution. The set of channels that ranked below 100 are (a) individual channels that users had to pay in “a la carte” billing such as pay-per-view and certain documentary and sports channels, (b) foreign news channels, e.g., Russia and France, which is

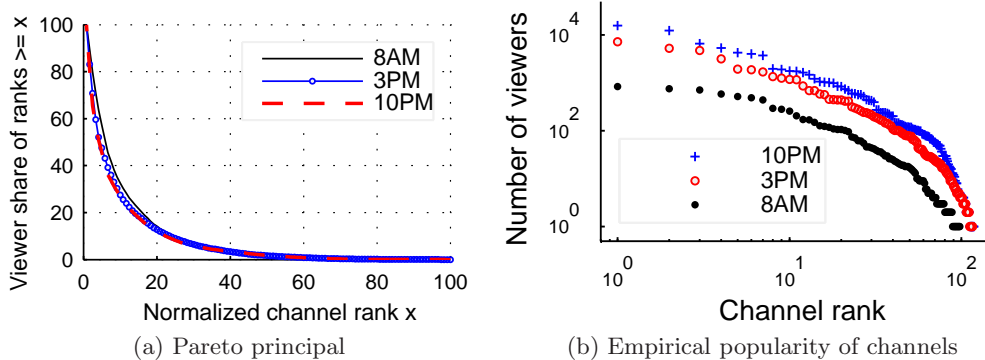


Figure 5: Channel popularity distribution

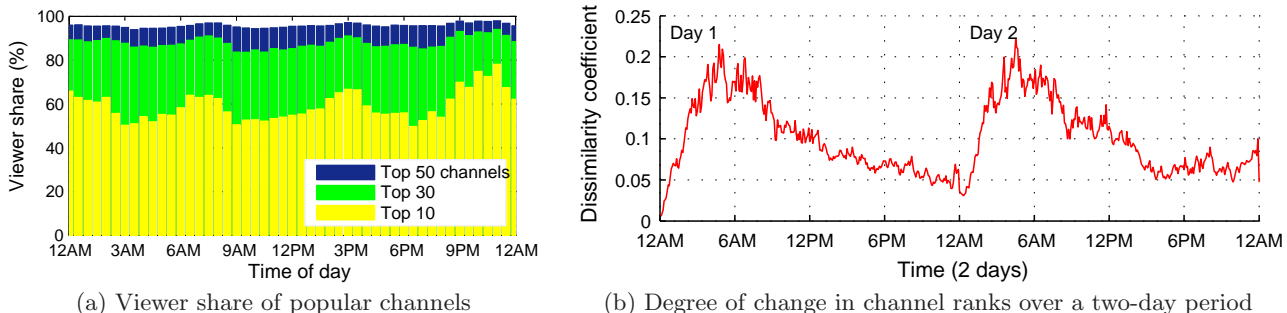


Figure 6: Channel popularity dynamics

spoken in a different language, (c) local channels of small regions, and (d) audio and reality shows.

4.2.2 Popularity dynamics

Next, we are interested in knowing to what extent channel ranks change over time. We examine the channel ranks for every 30 minute of a representative month period. At each time interval, we identify the groups of channels that make up the top 10, 30, and 50 ranks. We break ties in ranking by comparing the total viewing time of channels. We ran the analysis over the trace period and did not find differences between days. In Figure 6(a), we show the percentage of viewer share for the top ranked channels for a representative day. The average viewer shares are similar to that shown in Figure 5(a). However, the graph shows significant fluctuations across the day. Interestingly, popularity of top channels is reinforced around 8AM, 3PM, and 10PM. We conjecture this is due to the role of TV viewing as a “family” activity, where what to watch is no longer a single person’s decision but is shared amongst family members. (See [35] for a discussion of how group decisions are taken when watching television.) The user interests are more widely spread across channels during other times of the day.

Our next question is how frequent channel ranks change. Compared to the analysis on stability of popularity distribution in Figure 6(a), rank changes indicate the stability of popularity for individual channels. To measure the variations in channel ranks, we monitor the channel ranks for every 5-minute chunk of the trace period and calculate dis-

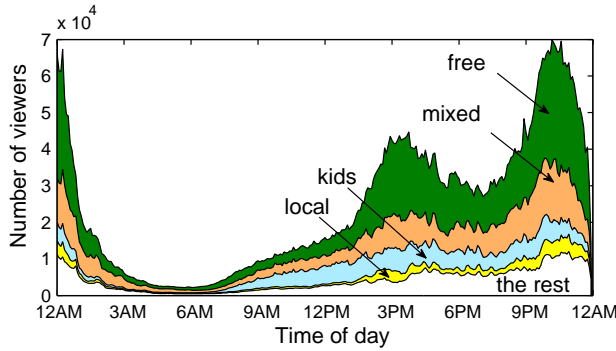
similarity coefficient for degree ranks $\xi = 1 - \rho^2$, where ρ is the Spearman’s rank-order correlation between channels of two time intervals. Spearman’s rank correlation compares the two set of values by converting each variable to ranks [36]. Kossinets and Watts, in their work to quantify the stability of node degree in social networks [37], also used Spearman rank-order correlation as a dissimilarity measure. In our case, the correlation is calculated,

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (1)$$

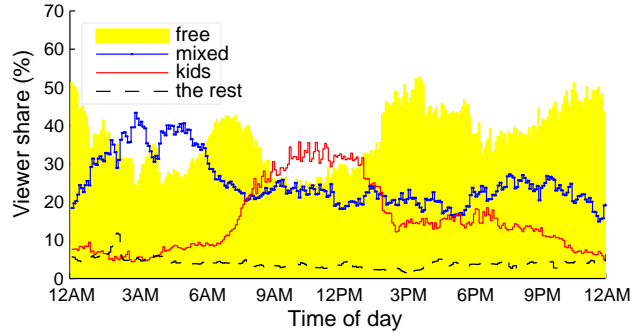
where d_i is the difference between ranks of the same channel at two time intervals, and n is the number of channels. The value of ξ varies between 0 and 1 and measures the proportion of variance in degree ranks that cannot be predicted from the initial set of ranks.

We compare the dissimilarity coefficient between channel rank snapshots of the first 5-minute time interval of Day 1 and the subsequent time intervals. We ran the analysis over the trace period and did not find differences between days. In Figure 6(b), we show the result for two representative days. Typically, a dissimilarity coefficient greater than 0.1 is considered to have substantial change in ranks.

Figure 6(b) shows periodic patterns over the two day period, indicating daily rank changes are repetitive. We expect scheduled programs to cause such periodic pattern. The dissimilarity coefficients increase at low usage hours, above 0.1. This indicates substantial degree of rank changes during those hours, supporting our finding in Figure 6(a). A



(a) The number of viewers per genre



(b) The percentage of viewer share per genre

Figure 7: Breakdown of viewers across genre over a representative day (Monday)

detailed look into the set of channels that compose the top popular list shows that 36, 63 and 90 distinct channels made the top 10, 30, and 50 ranks over the trace period. Only 1 channel consistently made the top 10 list. Our results suggest that while channel popularity distribution is skewed at any given moment, the set of channels that compose the most popular list changes dynamically over the course of a day.

One possible cause of the short-term focus on particular channels but long-term diversity of popular channels is that there are certain genres of programs that are periodically popular during certain times of the day or week. Also, one-time events may suddenly become popular for a short period of time, for instance, the Academy Awards. To examine the temporal correlation between popularity and genre, we show the aggregate number of viewers across content genres in Figure 7(a). We count the number of distinct viewers for every 10-minute time interval of a representative day. As expected, **free** content is the most popular throughout the day, except for the mid-day and around dawn when **kids** and **mixed** genres are the number one line up, respectively. While all other genres show diurnal pattern, the number of viewers for **kids** genre are relatively steady from early morning to late night. Figure 7(b) further shows the percentage of viewer share for **free**, **mixed**, and **kids**, and the remaining genres. The percentage of viewer share adds up to 100% at any moment and the most popular genre changes over a day.

4.3 Geographical locality

We now identify the content locality—the degree to which similar TV programs are watched by users in geographical proximity. We obtained information about geographical location of users from the IPTV provider. Naturally, we expect people to show more affinity to the **local** channels belonging to their own regions. However, it is not clear that the rest of channels will exhibit similar affinity across different geographical locations. To understand this, we compare the spatial locality of users’ interests at different levels of granularity (regions and DSLAMs) and over genres and channels.

4.3.1 Locality across regions

We first measure which content genres are popular across the 12 regions, which is a large subset of the whole country. Fig-

ure 8 shows the relative popularity of genres (based on the total time watched) for each region across a month period. The most popular genres are similar across regions: **free**, **mixed**, and **kids** channels are consistently popular. The correlation coefficient of the relative ranks of genres ranged from 0.97 to 0.99 between any two pairs of regions. Interestingly, users in some regions watched more **local** channels than those in other regions: viewing probability of **local** channels varied widely from 1% to 9%. This reflects that preference towards regional content is stronger in some regions than in others. Overall, the spatial variations in preference in genre across regions are up to 20%.

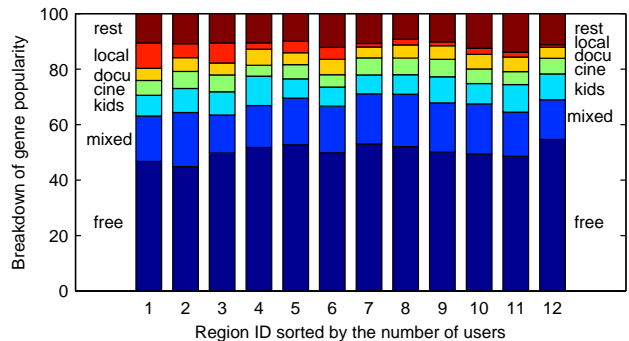


Figure 8: Popularity of genres across regions

4.3.2 Locality across DSLAMs

To analyze geographical locality at a smaller scale, we investigate how users in two different DSLAMs are similar in their viewing patterns. As a measure of DSLAM locality, for each channel and for all 670 DSLAMs, we calculated the ratio of the time spent on viewing of the corresponding channel against the total viewing time of all online users within the DSLAM. The locality value of 100% indicates homogeneous users: all online users at a given DSLAM watched the same channel. The locality value of 10% indicates that 10% of the viewing time of online users in the DSLAM was spent on the channel. By comparing the locality value across all DSLAMs, we may examine how users in different DSLAMs deviate from each other. Figure 9 shows the result for the top-ranked channel A, which is of **free** genre, based on a 9PM to 11PM trace of a representative day. Popularity of

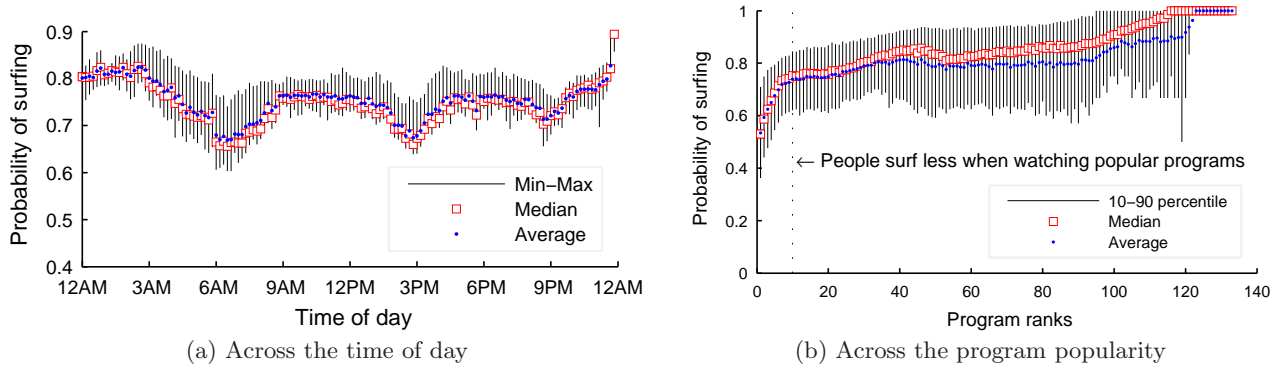


Figure 10: Probability of surfing varies across the time of day and the popularity of programs

the channel varied largely (10-42%) across DSLAMs. This means that 42% of total viewing time of the users was spent on channel A for one DSLAM, while for some DSLAMs only 10% of the total viewing time of online users was spent on channel A. Less popular channels showed smaller variations.

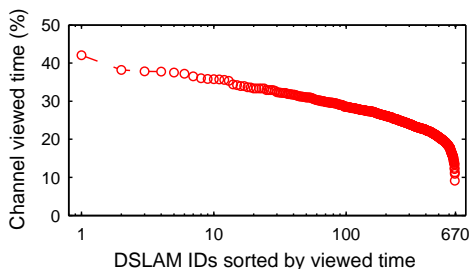


Figure 9: Popularity of a channel across DSLAMs

4.4 Summary

In this section, we have characterized the distribution of popularity of channels and studied the popularity evolution over time. Despite the fact that at any given point of time the popularity is highly skewed over a few channels, the set of top popular channels change over time, making the system highly dynamic. This may be explained by the fact that some genres are more popular at particular time of the day and that the audience base changes over time. We have also found that though certain channels are popular at any given time, the set of channels that are popular across regions and DSLAMs is not always the same. We have in fact found the variations up to 20% in the set of channels that people are watching across different regions or DSLAMs within the country.

5. CHANNEL SURFING PROBABILITY

In this section, we analyze channel switching behaviors to gain understanding of (a) what affects channel changes (e.g., time of the day, program popularity), (b) how people find interesting programs to watch based on the sequence of channels they visit, and (c) how people join and leave a particular channel.

5.1 Factors that affect channel changes

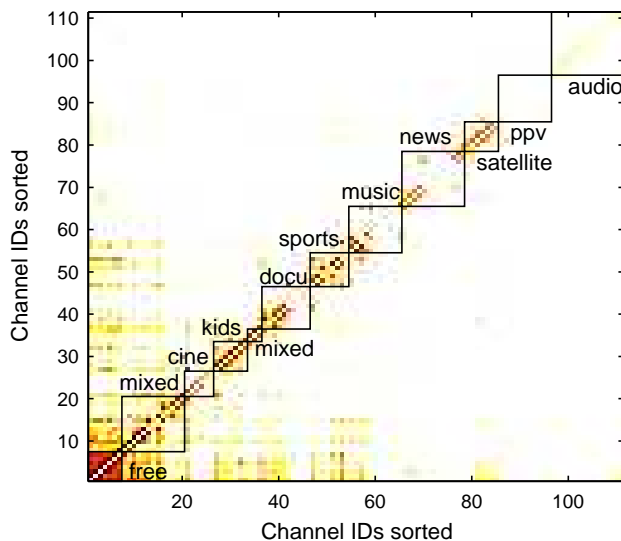
Table 1 shows that attention span of a user is typically genre-dependent and ranges from 5 to 21 minutes; genre clearly affects the likelihood and frequency of channel changes. In this section, we focus on other potential factors of channel changes: the time of day and program popularity. For these two factors, we calculate the probability that a given channel change is due to surfing compared to viewing in order to infer the user’s reason for channel switching. We call this the *surfing probability*, P_s . Let J be a set of channel join events during a given time interval. Then P_s is defined as follows:

$$P_s = \frac{|\text{surfing events in } J|}{|\text{surfing or viewing events in } J|}. \quad (2)$$

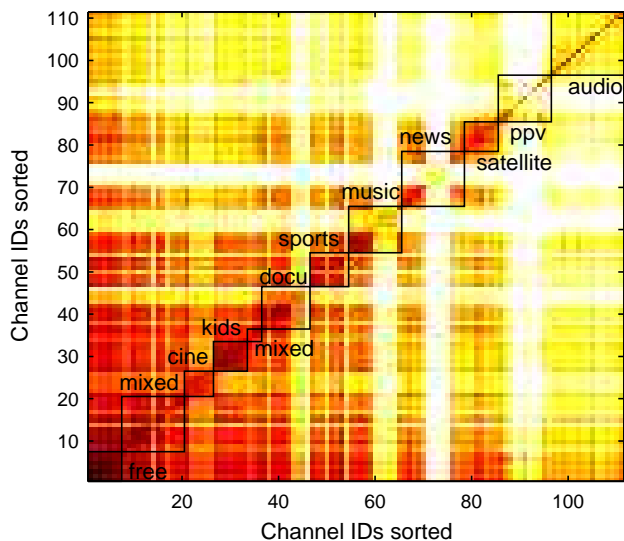
A surfing probability P_s of 0 indicates that after the channel change, the user will stay and view. While the probability of 1 means that channel change of a user will be followed by another channel change, indicating the user is surfing.

We first investigate how the surfing probability relates to the time of day. Figure 10(a) shows the minimum, the median, the average, and the maximum values of surfing probability for every 10 minutes of a month-long period. The median and the average values range from 0.65 to 0.90. The surfing probability decreases around 6-7AM, 3PM, and 9PM, which are the peak IPTV usage hours. In conjunction with our finding on the reinforced popularity of top ranked channels at peak times, the decrease in surfing probability at these hours indicates that a larger pool of concurrent viewers are more predictable and stable in their viewing patterns. The daily scheduled programs likely cause them to surf less during peak usage hours. The lowest point of surfing probability comes earlier than 8AM in the morning. We assume this phenomenon is related to long program length: three top ranked channels had a 3-hour long program from 6AM to 9AM. In contrast, we see increase in surfing probability in late evening and midnight, when the most popular programs are shorter in lengths (30 minutes to 1 hour) and when we expect a lot of people finish up some activity and check out what is on TV.

As a second potential factor, we investigate the influence of program popularity on channel changes. We compare the channel ranks and the surfing probability for every 10 minutes of a month period. Figure 10(b) shows the 10th percentile, the median, the average, and the 90th percentile



(a) Zapping probability from channel x to channel y



(b) Viewing probability from channel x to channel y

Figure 11: The probability of switching from channel x to channel y based on aggregate user behaviors: (a) channel change activities mostly reveal linear patterns; (b) viewing activities also reveal non-linear patterns

values of surfing probability against the program ranks. The surfing probability drops by 20% for 10% of the most popular programs. This indicates that viewers are more likely to stay tuned once they join a popular program. The correlation between the surfing probability and the program ranks is less obvious for medium to low ranked programs. For very unpopular programs (ranks beyond 100), we see increase in surfing probability. That is, users visiting unpopular channels are likely to switch channels immediately.

So far, we have used the viewing period of 1 minute to demarcate surfing instances from viewing instances. As given in Equation (2), the surfing probability is sensitive to the choice of viewing period. For validation, we repeated the three types of analysis in this subsection over different viewing periods: 10 seconds, 30 seconds, 2 minutes, and 5 minutes. We consistently found the same correlation between the surfing probability and the three potential factors.

5.2 Switching from one channel to another

With a TV remote control, people switch channels up and down easily in a sequential fashion. We call such channel changes *linear*. In IPTV, people may choose a different change strategy because there are simply too many channels and they can use an Electronic program guide (EPG). Using EPG, viewers can browse the broadcasting schedule overlaid on the screen and jump to arbitrary channels. Users also can directly input channel numbers through their TV remote controls and jump to any channel. The sequence of channels a user visits in such manner no longer needs to be sequential. We call the channel changes between a pair of channels that are not sequentially numbered *non-linear*.

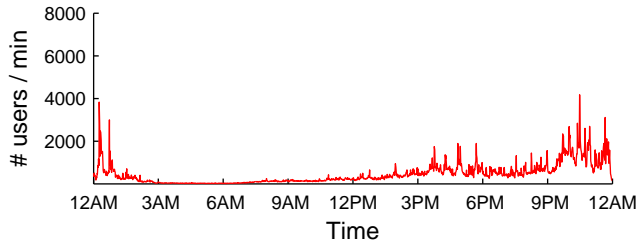
To understand the collective switching behaviors of all users, we examine the linear and non-linear switching patterns. Using a month-long trace, we calculate the normalized average probability of channel changes between every pair of channels and visualize them in Figure 11(a). While the actual channel numbers range from 0 to 265, we remove

gaps between channel numbers and reassign them sequentially in the matrix. A color pixel at (x, y) represents the probability of channel switching from channel x in the horizontal axis to channel y in the vertical axis. Darker pixel reflects higher probability. Since most channel changes are concentrated on a few pairs of channels, we use the log value of the probability to visualize both frequent and infrequent events. Therefore, any slight difference in the brightness of pixel is noteworthy.

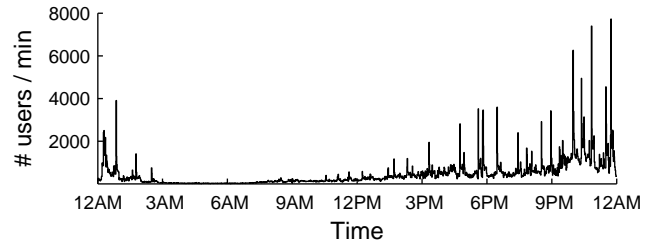
We observe several interesting channel switching habits. First, over 60% of channel changes are linear. The upward linear switches (on line $y = x + 1$) account for 43% and the downward linear switches ($y = x - 1$) account for 19%. Because channel changes happen between two different channels, the switching probability on line $y = x$ is zero. Second, certain genres show a distinctive pattern of non-linear channel changes within the genre, e.g., **free**, **sports**, and **kids**. This reflects that users are interested in the set of channels of similar topic. For example, a user tuning into one sports channel is likely to check out other sports channels, compared to movies or documentaries. In total, 76% of channel changes are within the same genre group, where 58% of switches are linear and 18% non-linear.

Third, the pattern of linear channel changes continues through the less popular channels like **music**, **satellite**, and **audio**. This indicates that linear channel changes, even with its inconvenience in IPTV compared to conventional TV, remain a characteristic behavior of users. Finally, the remaining 18% of channel changes are non-linear across different genres.

Next, we examine the influence of channel change patterns on viewing. Using the same trace, Figure 11(b) visualizes the probability of channel changes that led to viewing events, from channel x in the horizontal axis to channel y in the vertical axis. Intermediate surfing or away events are excluded in the analysis. We note that the sequence of two viewing events may not be relevant if they are separated by



(a) User arrival rate of channel A



(b) User departure rate of channel A

Figure 13: User arrival and departure evolution of the most popular free channel

away events. The results in this section were consistent even when we ignore viewing events separated by away events. As in Figure 11(a), a darker color represents logarithmically higher viewing probability. Compared to Figure 11(a), one striking difference is the consecutive viewing of the same channel on line $y = x$, which now accounts for 17% of all viewing instances. This reflects users' tendency to return to the same viewed channel for continued viewing, even after surfing. For example, viewers browse through other ongoing channels during commercials and rejoin the viewed channel. This may lead to missing some part of the programs, especially when users re-join few minutes late after the commercial break.

Another distinctive difference is the extent of non-linear viewing. Non-linear viewing patterns account for 67% of viewing instances, which is two times higher than that of switching probability. Non-linear viewing is increased for both inter-genre (40%) and intra-genre (27%) channel changes. The total intra-genre viewing instances account for 60%, indicating that viewers are inclined to watch programs of a similar topic. Within a given genre, some channels are viewed much less than others. This is due to the popularity differences of channels and the fact that some users are only subscribing to some channels (e.g., basic, family, a la carte). In summary, viewers tend to continue watching the same channel even after switching for some time and with high probability. This effect is called "inheritance" and is a well-established fact in traditional audience research (see [38], for instance).

5.3 User arrival and departure patterns

To understand the dynamics of user churn in the system, we analyze users' arrival and departure patterns of individual channels. We first demonstrate the arrival and departure rates for an example channel. We then investigate the inter-arrival and inter-departure times to infer the distribution of those processes.

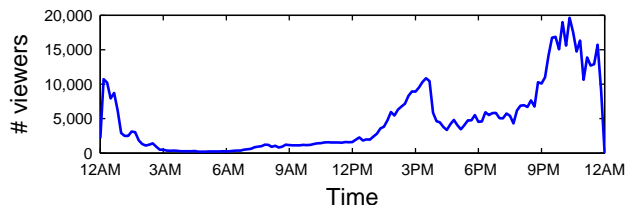


Figure 12: The number of viewers for channel A

5.3.1 Arrival and departure rates

Minute-by-minute user arrival and departure patterns demonstrate how users join and leave a particular channel. We discuss our findings on user arrival and departure processes using an example channel A, a top ranked **free** channel. Figure 13 shows the arrival and departure rates of users in channel A, over a representative day. The arrival and departure rates are similar on average. However, the latter ranges widely from 18 to 7,720 users per minute, while the former ranges from 18 to 4,172.

We make several observations. First, the arrival and departure rates vary over the day. A detailed analysis reveals that the variation is related to the viewer base. Figure 12 shows the number of concurrent viewers who watched channel A for longer than 1 minute in every 10 minutes of the day. The maximum values of user arrival and departure rates in Figure 13 follow the viewer base well. Our finding agrees with observation in other video systems [16, 33, 34].

Second, user departure patterns show consecutive spikes as in Figure 13(b), where the departure rate suddenly increases up to 5 times. These spikes are due to many users leaving the channel simultaneously at the end of programs or at advertisements. Hei *et al.* [34] also observed a batch departure pattern in a P2P-based IPTV system for a popular movie channel. However, they did not find such pattern in regular TV channels. Our IPTV system show spikes in departure patterns for popular channels across all genres. The difference between genres is the interval between spikes: **cin**e channels had a gap of 2 hours, while other channels showed shorter intervals between spikes. Some of the non-popular channels with few tens of concurrent users also show the batch departure patterns.

Third, the user arrival is much less time-correlated than the departure. Unlike Figure 13(b), the arrival patterns show much less variations. This indicates that the number of users watching a channel ramps up gradually over time, while it drops sharply at batch departure processes.

5.3.2 Inter arrival and departure times

Finally, we look at the inter arrival and departure times to identify the arrival and departure processes. Figure 14 shows the cumulative distribution of inter arrival and departure times for channel A, which is one of the top-ranked channels. The granularity of channel changes are 1 second in our trace and we assume arrivals are uniformly distributed within each second. The median inter arrival time is 0.07 seconds, indicating that $1/0.07 = 14$ users join channel A per second. The median inter departure time is also 0.07

seconds; users leave and join channel A at the same rate. The arrival rate varies over time and the arrival process is not stationary over the course of a day. The maximum inter arrival time is 14 seconds, which corresponds to the low usage hours. The maximum inter departure time is 700 seconds. While we do not show here, we find that both arrival and departure processes at short timescales (less than 1 millisecond) are exponentially distributed. This observation is consistent with previous studies on arrival processes of live streaming servers [33, 39, 40].

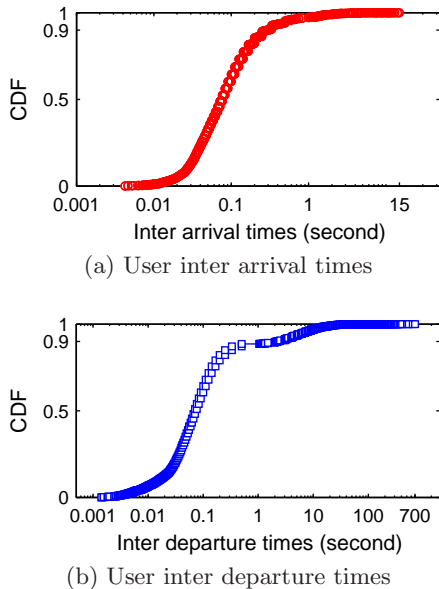


Figure 14: Inter arrival and departure time distribution of channel A

5.4 Summary

In this section, we have identified how people find content across three different angles. Understanding this mechanism is important because any user activity to search for content, imposes some load in the network. Surprisingly, we find that most of the channel changes do not lead to viewing activity, which highlights the fact that the network infrastructure could be overloaded with superfluous short channel surfing events. Our finding has major implications to IPTV networks, as opposed to traditional broadcast TV networks (e.g., cable and satellite TV). This is because in traditional broadcast networks, all channels are delivered to the end users and the surfing activities of users do not create additional load on the network.

We have also characterized the arrival and departure patterns of users of individual channels, where we have found time-correlated arrival and batch-like departure processes. Such dynamics in user churn put significant challenges to the design of P2P-based IPTV systems as they need to deal with drastic changes of the viewer population in a short period of time.

6. DISCUSSION

Our observations about IPTV workloads can help researchers to understand the user and the content charac-

teristics of TV and video streaming systems. Our findings also offer useful insights for the designers of existing and future video applications. For example:

- **Channel selection process:** Most channel changes are related to channel surfing and are short-intervalled. Therefore assisting users in channel selection is crucial to the quality of their viewing experience. In current IPTV and Internet video systems, electronic program guides (EPG) and metadata about channels such as tags and descriptions provide users with useful additional information about content. It is also important to make new content quickly available at the time of channel switching. This is more challenging for high-bandwidth video and TV streaming, especially in P2P distribution systems. For instance, the state of the art P2P Internet TV service Joost takes up to 5 seconds to switch channels. Another popular service, PPLive, buffers for 10 to 15 seconds before its payout [34]. Infrastructure-based Telco’s IPTV systems guarantee a switching delay of less than 1 second. However, this is not satisfactory as a 100-200 ms switching delay is considered “instantaneous” by viewers.

Our results about linear channel changes suggest that providing the channel numbers sorted for each user by either channel popularity for that user or dynamically based on current program popularity might reduce unnecessary channel changes. Our findings on the periodic popularity dynamics further suggest that a more dynamic re-numbering following time of day trends will be needed.

- **Popularity and program ratings:** An accurate measure of popularity is useful for TV viewers, content producers, and content distributors. The popular Nielsen ratings are usually aggregated over a long period of time and delayed a few hours to a day. With IPTV, we can get the instantaneous popularity information of programs and monitor its evolution closely. Having access to such realtime information is useful for TV viewers in deciding what program to watch.

- **Rewind and on-demand features:** Traditional and the first generation IPTV systems are live. Now TV is transforming itself to an interactive on-demand service and users are allowed to rewind back to any point back in time and watch TV at the time they want. Our findings on user arrival and departure patterns demonstrate that users’ arrivals to programs are less correlated than departures. Also users take a temporary leave during commercials and come back late, missing out on the beginning part of the programs they watch. For example, when watching a soccer game, users may miss a goal that happens immediately after a half-time break.

These behaviors call for supporting IPTV rewind functionalities, which have recently been offered by various operators under the name of “Past TV.” However, as opposed to DVRs which provide access to scenes stored locally, Past TV provides access to any scene regardless of whether the set-top-box was tuned into that channel or not. Under standard IPTV architectures, all rewind load would be directed to video servers. However, the strong group membership amongst users watching the same program reflects the potential for a peer-assisted IPTV rewind service, where latecomers in the program can easily find available seed peers. In our earlier work [32], we studied the efficacy of such peer-

assisted rewind service where users joining a channel always rewind back to the beginning of the on-going program. We found that with an assistance of a P2P system, the video server load dramatically reduces to 5%.

7. RELATED WORK

Understanding TV viewing habits in the old days relied on specialized monitoring boxes or phone surveys. Due to the difficulty of conducting a large-scale survey, the monitoring was limited to a subset of the TV viewing population. A representative example is the Nielsen Media Research [5] that provides a wide set of statistics on TV viewing and program ratings.

Viewing behaviors in Internet streaming have received recent attention. Sripanidkulchai *et al.* [33] studied the Internet radio and video streaming workloads of 5,000 users in Akamai, the largest content distribution network in the world. They observed a diurnal pattern in traffic and a Zipf-like popularity of live content. They found a heavy-tailed session duration and explained that short median session durations (less than 10 seconds) are likely due to channel surfing events. Our results are consistent with these earlier findings, but they are based on a specific IPTV system and the number of users is more than two orders of magnitude larger.

There are peer-to-peer Internet video streaming systems such as Joost, Zattoo, and BBC's iPlayer, that provide TV like viewing experience. Hei *et al.* [34] studied the packet traces from PPLive, a P2P-based IPTV system, and showed that peers join and leave at a higher rate at peak times. They observed a periodic (every 2 hour) batch user departure pattern for a popular movie channel. We also observed batch like departure pattern, however, at a smaller time scales (tens of minutes). This difference may be due to the channel switching delay of the two systems. Unlike our IPTV system, P2P-based IPTV has long start-up delays that range from tens of seconds to several minutes [34]. Due to the long playback lags users may behave somewhat differently from other TV viewing environments, for instance, users may be reluctant to switch channels during commercials. Other studies similarly analyzed the workloads and traffic characteristic of P2P-based IPTV systems [41, 42]. These results complement ours; while our system is infrastructure-based, understanding both systems provide a comprehensive view of IPTV workloads. In our earlier work [32], we looked into the potential for peer-assisted IPTV system.

Many previous studies have measured network properties of video-on-demand systems [16, 17, 34, 43, 44]. Cherkasova and Gupta [43] analyzed the enterprise media server workloads from Hewlett-Packard Corporation; Yu *et al.* [16] measured a commercial video system in China, PowerInfo; and Cheng *et al.* [17] analyzed the workloads of their high-bandwidth P2P-based VoD system. They discuss the popularity distribution of videos, viewing session characteristics, and their implications on designing efficient distribution architectures. Although our paper uses similar measurement techniques, the network environment we study is different as the viewing in live IPTV is not on-demand, but largely affected by the broadcaster's schedule, and ours is the first study is to use large-scale real traces from a commercial telco's IPTV service.

8. SUMMARY AND OUTLOOK

Television has long been a dominant and pervasive mass media and its viewing behavior has many important implications on the video distribution networks and systems. With the increasing popularity of Internet services, multi-media consumption patterns could change significantly over the coming years. However, as a recent report on the communications market stresses [22], the average time spent on watching television remains significant despite the uptake of broadband services. In fact, we expect TV to remain a crucial mass media in the future as it integrates many innovations in the Internet and as well as in the mobile space.

In this paper, we presented the first large-scale measurement study of a telco's IPTV system. We characterized several properties of viewing sessions, channel popularity dynamics, content locality, and channel switching behaviors. Our data analysis on the workloads of a live IPTV system brings values in understanding how the underlying networks and systems can better adapt to future challenges in Internet TV distribution.

There are several directions that we wish to pursue as future work. The first is to understand the implications of the workloads on the system design. Our trace showed that over 60% of a user's channel changes are related to channel surfing. We are interested in designing a system that can reduce the number of channels users need to sample before finding something that interests them. Second, we would like to study the profiles of individual users. While we limited our study to understanding the behavior of an average populace, future studies can focus on classifying users into different profiles (based on the time they watch TV, channels they watch, etc). Answering these questions are promising in the context of IPTV recommendation systems [45]. Finally, it would be very useful to develop a model that captures user switching behaviors and content popularity dynamics. The designers of future streaming systems can utilize such realistic models to test and assess their systems.

Acknowledgement

We thank Catherine Howell, Nuno Santos, Juan Antonio Navarro Pérez, Animesh Nandi, Nikos Laoutaris, and anonymous reviewers, for their valuable comments.

Sue Moon was supported by grant No. R01-2005-000-1112-0 from the Basic Research Program of KOSEF and by the IT R&D program of MKE/IITA [2007-S001-01,CASFI]. Xavier Amatriain was partially funded by an ICREA grant.

9. REFERENCES

- [1] Joost. <http://www.joost.com>.
- [2] Zattoo. <http://www.zattoo.com>.
- [3] Livestation. <http://www.livestation.com>.
- [4] BBC iPlayer. <http://www.bbc.co.uk/iplayer>.
- [5] The Nielsen Company. <http://www.nielsenmedia.com>.
- [6] J. G. Webster and L. W. Lichty, "Ratings Analysis: Theory and Practice," *The Public Opinion Quarterly*, vol. 56, no. 3, pp. 386-387, 1992.
- [7] IMS Research, "A Global Market Analysis, 2008 Edition," 2008.
- [8] Apple TV. <http://www.apple.com/appletv>.
- [9] Vudu. <http://www.vudu.com>.

- [10] Amazon Unbox. <http://www.amazon.com/gp/video/tivo/>.
- [11] S. E. Deering and D. R. Cheriton, "Multicast Routing in Datagram Internetworks and Extended LANs," *ACM TOCS*, 1990.
- [12] M. Cha, G. Choudhury, J. Yates, A. Shaikh, and S. Moon, "Case Study: Resilient Backbone Network Design for IPTV Services," in *Workshop on IPTV Services over WWW*, 2006.
- [13] AGB Nielsen Media Research, "Peplemeter." <http://www.agbnelsen.net/products/peplemeter.asp>.
- [14] T. Gitlin, *Inside Prime Time*. University of California Press, 2000.
- [15] M. Levy and S. Windhal, *Media Gratifications Research*, ch. The Concept of Audience Activity. Sage, 1985.
- [16] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng, "Understanding User Behavior in Large-Scale Video-on-demand Systems," in *ACM Eurosys*, 2006.
- [17] B. Cheng, L. Stein, H. Jin, and Z. Zheng, "Towards Cinematic Internet Video-On-Demand," in *ACM Eurosys*, 2008.
- [18] M. Dischinger, A. Haeberlen, K. P. Gummadi, and S. Saroiu., "Characterizing Residential Broadband Networks," in *ACM IMC*, 2007.
- [19] Office For National Statistics, United Kingdom, "Time Spent on Selected Free Time Activities of Full Time Workers," 2001.
- [20] Gary Holmes, Nielsen Media Research, "Nielsen Media Research Reports Television's Popularity Is Still Growing," September 2006.
- [21] Australian Film Commission, "Viewing Patterns: Hours of Viewing by Age Group," 2004. <http://www.afc.gov.au/gtp/wptvviewxage.html>.
- [22] Office of Communication, United Kingdom, "The Communications Market 2008," 2008. <http://www.ofcom.org.uk/research/cm/cmr08/tv/tv.pdf>.
- [23] Y. Ijiri and H. A. Simon, *Skew Distributions and the Size of Business Firms*. Elsevier North Holland, New York, 1997.
- [24] A.-L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5429, pp. 509–512, 1999.
- [25] H. Anderson, *Fixed Broadband Wireless System Design*. Wiley, 2003.
- [26] P. Reynolds, *Call Center Staffing*. The Call Center School Press, 2003.
- [27] A. Vazqueza, "Impact of Memory on Human Dynamics," *Physica A*, vol. 373, p. 747, 2007.
- [28] D. Brockmann, L. Hufnagel, and T. Geisel, "The Scaling Laws of Human Travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2005.
- [29] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms," in *IEEE INFOCOM*, 2006.
- [30] Statistics Canada, "Television Viewing," 2004. <http://www.statcan.ca/Daily/English/060331/d060331b.htm>.
- [31] S. Pedder, "Power in Your Hand," *The Economist*, April 2002.
- [32] M. Cha, P. Rodriguez, S. Moon, and J. Crowcroft, "On Next-Generation Telco-Managed P2P TV Architectures," in *International workshop on Peer-To-Peer Systems (IPTPS)*, 2008.
- [33] K. Sripanidkulchai, B. Maggs, and H. Zhang, "An Analysis of Live Streaming Workloads on the Internet," in *ACM IMC*, 2004.
- [34] X. Hei, C. Liang, J. Liang, Y. Liu, and K. Ross, "A Measurement Study of a Large-Scale P2P IPTV System," *IEEE Trans. on Multimedia*, vol. 9, no. 8, pp. 1672–1687, 2007.
- [35] J. Masthoff, "Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers," *User Modeling and User-Adapted Interaction*, vol. 14, pp. 37–85, 2004.
- [36] C. Spearman, "The Proof and Measurement of Association Between Two Things," *American Journal of Psychology*, vol. 100, no. 3/4, pp. 441–471, 1904.
- [37] G. Kossinets and D. J. Watts, "Empirical Analysis of an Evolving Social Network," *Science*, vol. 311, no. 5757, pp. 88–90, 2006.
- [38] J. G. Webster, "Audience Flow Past and Present: Television Inheritance Effects Reconsidered," *Journal of Broadcasting & Electronic Media*, 2006.
- [39] E. Veloso, V. Almeida, W. Meira, A. Bestavros, and S. Jin, "A Hierarchical Characterization of a Live Streaming Media Workload," in *IMW*, 2002.
- [40] K. C. Almeroth and M. H. Ammar, "Collecting and Modeling the Join/Leave Behavior of Multicast Group Members in the MBone," in *International Symposium on High Performance Distributed Computing (HPDC)*, 1996.
- [41] X. Zhang, J. Liu, B. Li, and T.-S. P. Yum, "CoolStreaming/ DONet: A Data-driven Overlay Network for Peer-to-Peer Live Media Streaming," in *IEEE INFOCOM*, 2005.
- [42] T. Silverston and O. Fourmaux, "Measuring P2P IPTV Systems," in *NOSSDAV*, 2007.
- [43] L. Cherkasova and M. Gupta, "Analysis of Enterprise Media Server Workloads: Access Patterns, Locality, Content Evolution, and Rates of Change," *IEEE/ACM ToN*, vol. 12, pp. 781–794, 2004.
- [44] C. Huang, J. Li, and K. W. Ross, "Can Internet VoD be Profitable?," in *ACM SIGCOMM*, 2007.
- [45] L. Ardissono, C. Gena, P. Torasso, F. Bellifemine, A. Difino, and B. Negro, *Personalized Digital Television. Targeting Programs to Individual Users*, ch. User Modeling and Recommendation Techniques for Personalized Electronic Program Guides. Kluwer, 2004.