POLITECNICO DI TORINO
Repository ISTITUZIONALE

Passive characterization of sopcast usage in residential ISPs

Original
Passive characterization of sopcast usage in residential ISPs / BERMUDEZ CORRALES, I.N., Mellia, M., Meo, M.. - STAMPA. - (2011), pp. 1-9. (IEEE P2P 2011 Kyoto, Japan August 2011) [10.1109/P2P.2011.6038655].

Availability:
This version is available at: 11583/2495551 since:

Publisher:
IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, 445 HOES LANE, PISCATAWAY, NJ 08855 USA

Published
DOI:10.1109/P2P.2011.6038655

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

09 October 2026

Passive Characterization of SopCast Usage in Residential ISPs

Ignacio Bermudez
Politecnico di Torino, Italy
ignacio.bermudez@polito.it

Marco Mellia
Politecnico di Torino, Italy
marco.mellia@polito.it

Michela Meo
Politecnico di Torino, Italy
michela.meo@polito.it

***Abstract*—In this paper we present an extensive analysis of traffic generated by SopCast users and collected from operative networks of three national ISPs in Europe. After more than a year of continuous monitoring, we present results about the popularity of SopCast which is the largely preferred application in the studied networks. We focus on analysis of (i) application and bandwidth usage at different time scales, (ii) peer lifetime, arrival and departure processes, (iii) peer localization in the world.**

Results provide useful insights into users' behavior, including their attitude towards P2P-TV application usage and the consequent generated load on the network, that is quite variable based on the access technology and geographical location. Our findings are interesting to Researchers interested in the investigation of users' attitude towards P2P-TV services, to foresee new trends in the future usage of the Internet, and to augment the design of their application.

I. Introduction

In the recent years we have witnessed the success of P2P-TV applications, bringing TV channels, some of which live, to the users' home through the Internet. Several commercial P2P-TV systems such as SopCast, PPLive, TVAnts, among the most widespread ones, are already available and pretty much popular among users because they feature cheaper video broadcasting than other solutions, e.g., IPTV or pay-TV. P2P-TV traffic characterization has thus become a topic of great interest for the research community [1], [2].

Service providers, network operators and designers, are interested in assessing the potential impact of this traffic on the network of today, impact that might turn out to be disruptive, given the possible large number of users and high bandwidth requirement combined with the traffic being loosely controlled with respect to network conditions. Researchers are interested in the investigation of users' attitude towards these new services to foresee new trends in the future usage of the Internet, and to augment the design of their application. A deep understanding of P2P-TV traffic and its characterization is therefore an important task that can contribute to the design of network elements, including traffic engineering mechanisms, component dimensioning, resource management strategies.

In this work, we contribute to the characterization of P2P-TV traffic by analyzing the traffic due to popular applications (SopCast, TV-Ants and PPLive), in the operative links of four networks in operation in Europe, three of which provide ADSL access, the forth one employs FTTH (Fiber-To-The-Home) technology. Differently from the measurement works present in the literature, we adopt a pure passive methodology to observe normal usage of P2P-TV applications by customers. Collecting traffic for more than one year, we found that SopCast is the largely preferred application by customers in these networks. Furthermore, the usage of these applications is still very much discontinuous and often associated to events, such as sport events, that are popular but expensive to retrieve through normal TV broadcasting systems. We then focus on two months during which the UEFA Champions League 2009 final matches were held. Investigating deeper into the SopCast traces, we observe traffic and peer volumes, swarm evolution, peers' geo-localization and lifetime, and their contribution to the video distribution. Results suggest that the implications of traffic burstiness, the peer population and their evolution might become challenging for the network, should these applications become widely popular.

The results presented in this paper allow to highlight some key aspects of the usage of P2P-TV systems by European users:

- Even though the daily bandwidth usage of P2P-TV applications is not significant, it can be substantial during periods in which popular events are shown. Today few tens of users already contribute upto 15% of total aggregate traffic generated by 20.000 customers.
- Node churning during the lifetime of a stream is not significant, but there is a flash crowd entering the system at the beginning of the event and a rush towards exit at the end. This clearly has an impact on the design of P2P-TV applications.
- Evidence shows that it is often high-speed residential networks (FTTH) and University networks that altruistically serve content to residential peers with highly asymmetric bandwidth. Without the contribution of those peers, the P2P-TV system would not sustain the service.
- Geo-locality of swarms formed around distributing video from different channels is deeply affected by cultural and language trait of customers.

The latter two facts clearly impact the ability to localize P2P traffic, a theme that is currently debated in the research community. We then discuss their implications in the case of P2P-TV systems.

II. Related work

The interest in understanding P2P streaming applications has raised in the last years. This is due to both the increasing

TABLE I
PROBES CHARACTERISTICS

Name	Cust.	Technology	CC
TP	10k	ADSL 0.5/6Mb/s	PL
IT-ADSL	15k	ADSL 1/20Mb/s	IT
IT-FTTH	4k	FTTH 10/10Mb/s	IT
MT	4k	ADSL 0.5/5Mb/s	HU

interest about P2P technologies in science, and on the large available solutions, most of which adopt proprietary and unknown solutions. This justifies some researches, as [3]–[5], who have focused on single commercial system, and investigated their internals using active methodologies. Other works instead study aspects regarding to P2P streaming systems by observing the behavior of some peers in local testbed. [6] investigates the stability of nodes in PPLive, [7] focuses in on the node degrees of popular versus unpopular channels in PPLive. Alternatively, [8], [9] measure issues related to P2P Quality of Experience offered to users. Results presented in [8] are focused in inferring metrics such as chunk propagation delay, start-up latency, network-wide playback continuity and playback lags among peers by actively crawling PPLive buffer maps, instead [10] focuses on audio-video synchronization, TV channel zapping time, blocking probability, etc. Authors exploit logs made available from unspecified commercial P2P streaming system.

[11] and [12] present statistics collected from a large-scale live event broadcasted on P2P networks by PPLive, PPStream, SopCast and TVAnts. Their analysis is limited by their experimental framework in which few peers participate in the P2P overlay. Similarly, [2] provides similar and complementary measurements, again observing the behavior of 40 peers running in 5 countries in Europe. Finally, in [2] authors provide some insights in the peer selection system adopted by popular P2P-TV applications. Again, artificial testbed were used.

In this paper we instead present results collected passively monitoring actual users, running the application at home at their willing. We do not have control on any peer, and we do not alter the P2P-TV system under observation. Having access to several probes, we characterize the typical usage of P2P-TV system, and the impact they have on the network.

III. EXPERIMENTAL SETUP

Our work is based on the data collected during monitoring experiments performed in the context of the Network Aware Peer-to-Peer Application under WIse NEtwork (NAPA-WINE Project), funded by the EU in the Seventh Framework Programme [13]. Several traffic monitoring probes were installed to passively collect packet level traces from ISP network operational links. Traces were collected by running the Tstat [14] traffic analyzer on each probe machine. Through a Deep Packet Inspection (DPI) technique, Tstat is instructed to identify P2P-TV traffic of prominent and popular commercial P2P-TV systems, namely TV-Ants, PPLive and SopCast. Packets belonging to those applications are dumped on output files to be later post-processed. DPI rules have been manually tuned and verified using both laboratories testbed, and experiments in the wild, showing very reliable results in terms of both true and false positives [15].

Probes are located on aggregation points (Point-of-Presence, PoP) of three European Internet service providers (ISPs). Each vantage point monitors thousands of residential users, accessing the network via DSL or FTTH lines. The main characteristics of the 4 probes are summarized in Tab. I, which reports the name used throughout the paper, the approximate number of aggregated customers, the offered access technology, maximum upload/download capacity offered to users and the country (CC) the probe is placed in. As it can be observed, the set of probes is very heterogeneous: they span over three different countries, using both ADSL, FTTH access technologies. Considering the access capacity, ADSL technology offers the users different bitrate depending on the type of contract with the ISP and on the quality of the physical medium, ranging from 2 to 20 Mb/s downstream and up to 1024 kb/s upstream. IT-FTTH users enjoy 10 Mb/s Ethernet based full-duplex connectivity. IT-ADSL and IT-FTTH probes are in the same ISP in Italy.

In the following we refer to a "user" as the person that is using the P2P-TV application, while a "peer" refers to the application running and exchanging traffic. Peers and users are uniquely identified by their IP address. We distinguish between *internal* peers, i.e., the peers ran by users inside the monitored PoP; and *external* peers, i.e., those peers ran by Internet users. Similarly, we then define *incoming* traffic (RX) the one flowing from external peers to internal peers, and *outgoing* traffic (TX) the one flowing in the opposite direction. In addition, we use the popular term "*swarm*" to refer to a set of peers which are connected in a P2P manner to watch the same TV channel.

IV. CHARACTERIZATION OF THE P2P-TV USAGE

We start by characterizing the usage and popularity of P2P-TV among users inside the PoPs. Fig. 1 reports the P2P-TV average daily bitrate observed at the TP vantage point. Top plot refers to a one year long period of time, starting from the 15th of February 2009; a zoom in a one-week-long period of time during April 2010 is shown in bottom plot.

On average, the traffic generated by these applications is marginal, accounting to at most 5% of the average volume of traffic seen in the PoP. However, the burstiness of traffic, better visible in the weekly plot, shows that P2P-TV usage is concentrated during short periods of time in which the amount of generated traffic can reach very high and possibly disruptive peaks. We observe that the volume of traffic generated by P2P-TV typically coincides with the transmission of popular sport events, e.g., UEFA Champions League during Wednesday and Thursday or Premier League (England First Division) on Saturday and Sunday. The amount of traffic due to P2P-TV applications during those events often exceeds 15% of total traffic in the PoP, more than the whole YouTube traffic observed at the same time in the same vantage point. This "bursty" user behavior, which can be pretty difficult to handle, is very different from TV and IPTV users' behavior, whose

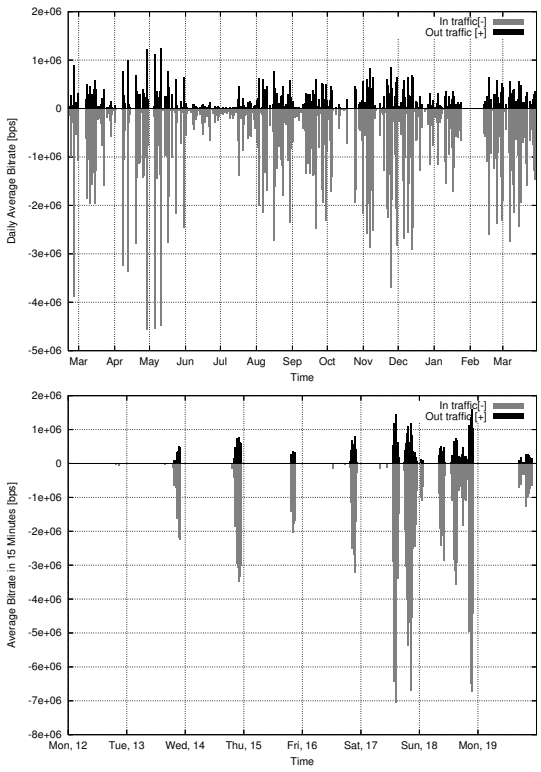


Fig. 1. Average bitrate of P2P-TV traffic collected in TP PoP.

access to the service is smoother and more evenly distributed over time.

Considering the IT-ADSL, IT-FTTH and MT probes, we see the identical spotty behavior, but the fraction of P2P-TV traffic is in general smaller: it never exceeds 3% of total traffic, showing that the P2P-TV applications are less popular in those countries.

Considering the popularity of the three applications, SopCast is by large the most popular one. During April 2009, PPLive accounted for about 3% of total P2P-TV traffic, while TVAnts usage barely reached 0.5%. This holds true in all monitored networks. For the sake of correctness, we verified that the Tstat DPI was correctly able to identify PPLive and TVAnts traffic by running some peers in our campus network. Results were positive. Therefore, we can conclude that in the monitored networks, SopCast is much more popular than the other two P2P-TV systems.

In the following, we restrict our analysis to the largest traces of SopCast traffic. Table II summarizes the subset of selected traces. They all correspond to popular sport events that were held from April to May 2009 when the semifinal and final of UEFA Champions League was held. The table reports also the number of internal and external peers and the total RX and TX traffic. The largest trace corresponds to the UEFA semifinal match held on April 28^{th}. 177 different internal peers were present, downloading more than 38GB of traffic in less than 4 hours, on the other hand only 7GB of traffic was uploaded.

TABLE II
MAIN SET OF TRACES

Date	Time	Length	Int. Peers	Ext. Peers	RX	TX
11 Apr	10:50	7 Hours	62	101418	17GB	4GB
11 Apr	17:50	4 Hours	48	33536	12GB	3GB
18 Apr	15:00	8 Hours	71	110645	27GB	4GB
21 Apr	16:30	5 Hours	47	80013	9GB	2GB
28 Apr	17:55	4 Hours	177	160181	38GB	7GB
5 May	18:15	3 Hours	133	149826	37GB	6GB
6 May	17:15	4 Hours	40	42264	10GB	2GB
10 May	13:30	5 Hours	101	96279	5GB	1GB
19 May	17:15	3 Hours	31	20872	8GB	2GB
27 May	17:45	4 Hours	19	18619	6GB	1GB
30 May	15:00	3 Hours	44	35599	15GB	3GB

V. SWARMING ANALYSIS

The information on whether the observed peers are watching the same "*channel*" [1] is interesting since it can be leveraged to better characterize the user's behavior. Unfortunately, identifying the channel turns out to be complex from passive monitoring of uncontrolled peers. Moreover SopCast adopts proprietary protocols and uses encryption mechanisms, that make harder to get done with the channel identification.

To avoid the complex (and questionable) reverse engineering of the SopCast protocol, we define a methodology that allows to group peers in swarms. This methodology is generic and can be leveraged for other systems as well. The intuition at the base of our solution is that peers with similar neighborhoods[2] are probably belonging to the same swarm; then, if peer a and peer b have a lot of neighbors in common, we claim that they are watching the same channel. On the contrary, if peer a and peer c have only a few neighbors in common, we claim they belong to two different swarms.

Let a and b denote two internal peers and let $P(a)$ be the set of peers *contacted* by a, i.e., peers which a sent a packet to. The amount of common peers among a and b is then $C(a,b) = |P(a) \cap P(b)|$, where $|\cdot|$ is the cardinality operator. We define then the common peer matrix M, as a matrix in which element (i,j) is $M_{ij} = C(i,j)$.

By sorting peers so that two adjacent rows (columns) in M refer to peers that have a large number of common peers we can then easily identify the swarms. For the peer sorting, we use the following measure. Let V_a be the vector of common peers of a with all other peers, i.e., the a-th row of M. Denote by V_a^T the transposed of V_a, i.e., the a-th column of M. The product,

$$S(a,b) = 2\frac{V_a V_b^T}{V_a V_a^T + V_b V_b^T} \quad (1)$$

is a measure of the similarity between the neighborhoods of a and b. By iteratively sorting the list of peers and moving closer those with larger similarity, we obtain the *swarming matrix*, i.e., an ordered common peer matrix M' that depicts in a clear way how peers are grouped together.

Figure 2 reports the results obtained considering the 133 peers that were active for more than 600 s on the 5^{th} of

[1] Peers watching different channels belong to different disjoint swarms

[2] A peer neighborhood is the set of peers it exchanges data with.

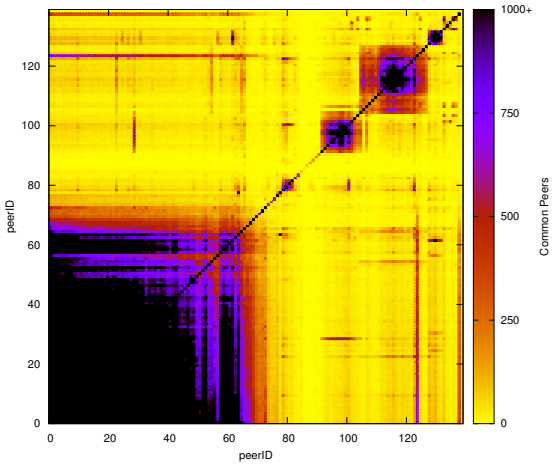


Fig. 2. Swarming matrix for 5th of May trace. Darker blocks refer to peers in the same swarm.

May. Each cell is colored according to the amount of common peers it represents. The numbers along the main diagonal correspond to the total number of contacted peers, $P(a)$. The swarming matrix shows that there are several groups of peers that share a large fraction of common peers, identified by the darker blocks. The largest block includes peers from 0 to 70 (that we name swarm A), the second group corresponds to peers from 105 to 126 (swarm B), then peers from 90 and 105 (swarm C). The magnitude of well defined groups of peers with intersections of neighborhood suggests that they correspond to different swarms, or channels. We claim that during the 5^{th} of May event users were watching different channels that were possibly broadcasting the same event. As we will see later, each identified swarm is also characterized by very different properties, which corroborate this claim.

Some common peers between different groups are also visible, corresponding to peers that changed swarm (channel) during the considered period. For example, peer 123 has a large number of common peers with both swarm A and swarm B. Let X and Y be the sets of internal peers that belong to different swarms. Then,

$$P(X, \bar{Y}) = \cup_{x \in X} P(x) \setminus \cup_{y \in Y} P(y) \tag{2}$$

is the set of unique external peers contacted by peers of swarm X but not by peers of swarm Y. Let $P_{\Delta T}(a)$ be the set of peers contacted by the internal peer a during time ΔT. We define the *swarm affinity* of peer a to swarm X and not to Y as

$$A_{\Delta T}(a, X, \bar{Y}) = 100 \frac{|P(X, \bar{Y}) \cap P_{\Delta T}(a)|}{|P_{\Delta T}(a)|} \tag{3}$$

Fig. 3 shows the affinity of peer 123 to swarm A and swarm B, considering $\Delta T = 5$ minutes moving along the event duration. The plot shows that peer 123 exhibits a high swarm affinity towards swarm B from 18:30 until 20:10, time at which its affinity to swarm B drops and the one to swarm A increases. We claim that peer 123 left swarm B to join swarm A at 20:10. Identical results are obtained when considering other cases.

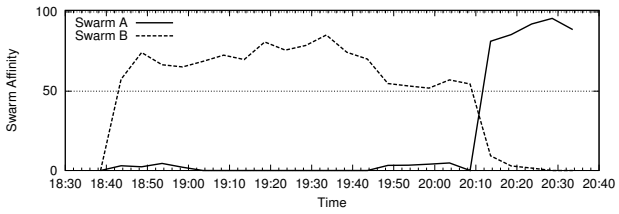


Fig. 3. Affinity of peer 123 to swarm A and B during the 5th of May trace.

Swarming matrices allow us to group users watching the same channel. As a first result, we can clearly see that some channels are more popular in the monitored network. SopCast usually provides more than one channel for the same transmitted event, so channel zapping might corresponds to users looking for better channel features such as video quality, sound quality, channel stability or even language.

We repeated the swarm analysis on all traces, identifying several swarms. In the following, we restrict our analysis to the subset of swarms reported in Table III, which are the largest swarms in terms of number of peers. The Table summarizes prominent swarm characteristics: the number of internal and external peers, the total amount of received (RX) and transmitted (TX) data, estimated video rate, probe country code (CC) location and the portion of external peers that belong to the same Autonomous System (AS) the probe was located in. Note that channels are sorted by decreasing values of the last metric. Note that all the largest swarms were observed in the TP traces, being P2P-TV usage more popular in Poland than in the other two European countries. Nonetheless, in swarm 11 and swarm 14 we identified one peer that was monitored in MT and IT-ADSL, respectively. These are listed in the two bottom rows of the Table.

The estimated video rate, is computed considering the number of peers watching the channel and the total amount of video data observed, during a window time of 1 minute. The video data is discriminated from the signaling data by taking advantage of the very biased distribution of packet size of SopCast. Furthermore packets which hold more than 1000 bytes in their payloads are considered as video packets. Results show that the video rate is typically lower than 480kbps, i.e., low quality video.

VI. Users behavior

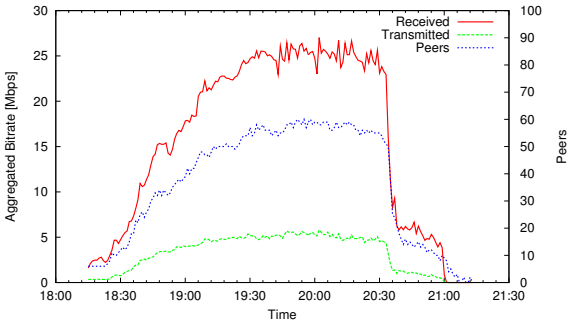
In this section we focus on the user behavior, and in particular, on the users' lifetime, arrival and departure process, and the traffic they receive and send during their activity time.

A. Evolution versus time

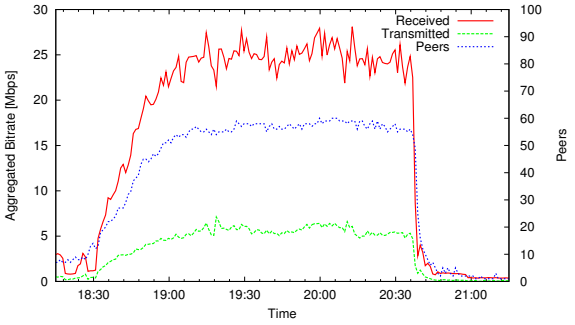
Fig. 4 reports the amount of active peers (right y-axis, blue dotted lines) and the aggregate bitrate they exchange (RX, left y-axis, solid red line; TX, left y-axis, green dotted line) versus time at sampling rate. Figs. 4(a) and 4(b) refer to Swarm 5 and 6, respectively; these swarms were observed in TP on the two days of the UEFA Champions League semi-final matches; matches started at 18:30 GMT+0 and finished

TABLE III
LIST OF THE LARGEST SWARMS

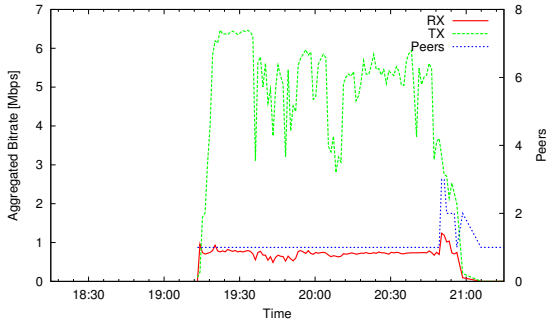
Swarm ID	Internal Peers	External Peers	RX [GB]	TX [GB]	Video Rate [kbps]	CC	PL AS %
0	35	15489	8	2	330	PL	32
1	29	19701	8	2	400	PL	31
2	50	32757	15	3	450	PL	28
3	33	25575	9	2	400	PL	27
4	41	33320	15	3	400	PL	27
5	69	60502	23	5	420	PL	25
6	66	76416	24	5	470	PL	23
7	19	20662	8	2	350	PL	21
8	77	68264	30	6	400	PL	19
9	11	10371	3	1	440	PL	10
10	5	12288	1.5	0.2	320	PL	9
11	12	18684	5	1	430	PL	9
12	10	20930	2	0.2	370	PL	8
13	8	13591	2.6	0.4	450	PL	8
14	16	39948	4.3	0.4	380	PL	6
15	13	39718	4	0.3	450	PL	5
16	25	54830	5	1	470	PL	4
17	8	23333	2.7	0.3	480	PL	3
18	10	30026	3	0.5	330	PL	2
19	9	27195	2.1	0.5	400	PL	2
11	1	8049	0.5	0.2	430	HU	N/A
14	1	5122	0.3	0.1	380	IT	N/A



(a) Swarm 5 evolution in TP



(b) Swarm 6 evolution in TP



(c) Swarm 14 evolution in IT-FTTH

Fig. 4. Number of peers, amount of aggregated transmitted and received bitrate in 3 swarms.

at 20:35 GMT+0. Fig. 4(c) refers to Swarm 14 and it was collected in the IT-FTTH probe on May 19th, during the "Copa del Rey" match held in Spain.

Results show a very synchronized and impulsive behavior of users supporting that the use of P2P-TV applications is more linked to sport events than to normal TV broadcasting. Most peers arrive at match starting time (18:30) and leave when the match ends (20:35).

Considering Figs. 4(a) and 4(b) which refer to ADSL customers, the incoming bitrate reaches 25 Mb/s, corresponding to more than 4% of total incoming traffic in the PoP for a just single swarm. Notice that, despite the large averaging periods of 1 s and the almost stable number of peers during the second half of the event, the variability of the incoming traffic is significantly high. This hints to a very high burstiness in the packet arrival process the network has to handle. For outgoing traffic, it is interesting to notice that the total amount of traffic the ADSL peers can contribute to transmit is limited to 1/5 of the traffic they receive, meaning that internal peers marginally contribute to the P2P-TV data exchange. This holds true for all swarms in Tab. III. On the contrary, the (only) active peer in IT-FTTH PoP in Fig. 4(c) exploits the much larger upload capacity offered by the FTTH access technology. This single peer transmits at [5,6]Mb/s, i.e., 5 times larger than the received rate, in this way acting as a "*super peer*". This phenomenon is documented by SopCast engineers which claim a super peer can transmit at most five copies of the data it downloads. Also in this case, traffic burstiness is very large.

B. *Channel popularity*

Fig. 5 shows the peer evolution for three different contemporary swarms in the TP probe. Swarm 5 is much more popular in the monitored PoP, reaching 60 coexisting internal peers (left y-axis); swarms 10 and 14 are much less popular (right y-axis). Nonetheless, the peers evolution is very similar, suggesting that users are watching different channels, but the same event. Recall indeed that SopCast (and P2P-TV systems in general) typically offers several channels that broadcast the same event.

We are now interested to observe if there is any bias in the channel popularity that can be related to the cultural traits of a given county. Let P_l be "local channel popularity", i.e., the fraction of internal peers watching a specific channel over all internal peers that were alive during an event. Let P_c be "Polish channel popularity",i.e., the the fraction of peers in

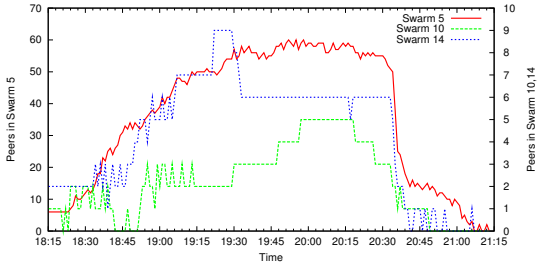


Fig. 5. Peer evolution for three swarms present at the same time.

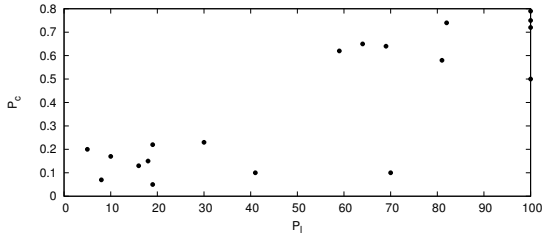


Fig. 6. TP Popularity versus Country Popularity and Video Rate

Poland over the entire peer population that joined a channel. P_l is a measure of how popular is a channel in the vantage point (in Poland). P_c instead is a measure of how popular a channel is among Polish with respect to worldwide population, i.e., P_c measures how biased the peer distribution of a channel is towards Poland.

Fig. 6 shows P_c versus P_l for each channel. Interestingly there are two main clusters of points: channels which are locally popular (large P_l) and mostly popular in Poland (large P_c), and channels which are not locally popular (small P_l) but popular worldwide (small P_c). The first subset corresponds to channels that exhibit a high bias toward Polish interests so that they are mainly popular in Poland. The second subset on the contrary corresponds to channels which are less interesting for Polish and more popular outside Poland. This hints that cultural and language traits affect the channel selection. Interestingly, there is one channel with high P_l but low P_c (local channel popularity of 70%, Polish channel popularity 10%). This particular channel results globally interesting, so that the fraction of Polish over all peers is not predominant despite the large interest of Polish toward the content (indeed the event being broadcast at that time was the match of Liverpool v.s. Arsenal in April 21^{th} 2009). In Section VII-A we will elaborate further on this finding.

C. Peer arrival and departure processes

To investigate further the users' habits in joining a P2P-TV swarm, Fig. 7 depicts the time when a user, labeled by an identification number, starts the application and the time it stops it; the horizontal segment represents the time the user is watching the event using SopCast; we denote it as the "*lifetime*" of users. Results show that users start the application some minutes before or after the match start, but once the event finishes they stop it in a synchronized way. To obtain more granular results about user arrival process and lifetime. Fig. 8

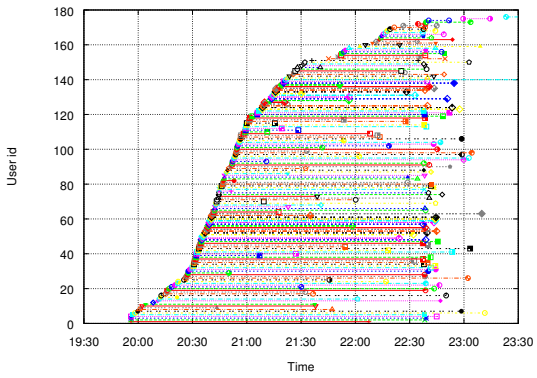


Fig. 7. TP internal peers lifetime.

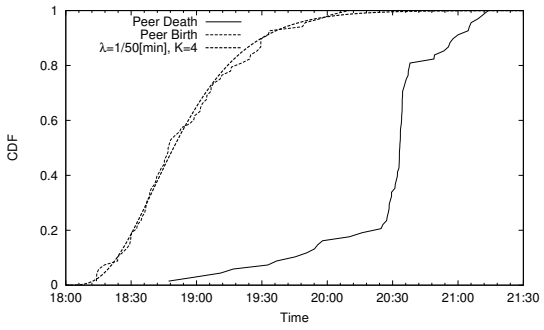


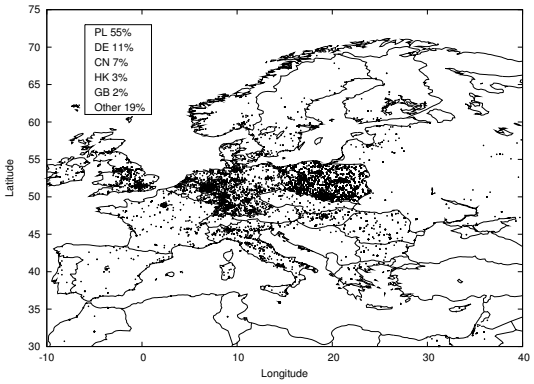
Fig. 8. Peer arrival and departure processes in a popular swarm.

illustrates the Cumulative Distribution Functions for the arrival and departure process of peers for a single channel . The arrival process CDF fits well an Erlang CDF with shape parameter $K = 4$ as reported in the figure. This hints to an arrival process variance much smaller than a Poisson arrival process, highlighting the users' synchronization with the event starting time. Even more clear is the "sudden death" phenomenon that is observed at the end of the event. Considering departures before 20:30, 20% of peers have already left the channel at almost constant rate of 2% peers/minute due to intrinsic channel churning.

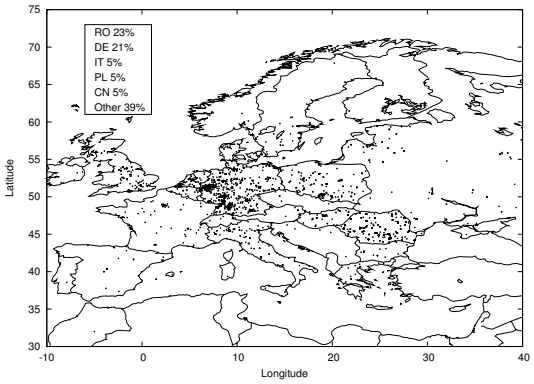
Recalling that peers rarely change channel during their lifetime, we conclude that in general node churning is not significant during an event. However, we believe that P2P-TV designers have to cope with the flash crowd entering the system at the beginning of the event and a rush towards exit at the end. Both events indeed stress the system that observes a sudden increase/decrease of peers and resources.

VII. Spatial analysis

We now focus on the spatial characteristics of external peers investigating whether is there any localization mechanism that drives peer discovery and selection process, or any cultural bias that influences the P2P-TV overall distribution characteristics. Each observed external peer has been geographically mapped using information provided by MaxMind GeoIP lite databases [16] which are coherent in date to the traces used.



(a) TP



(b) IT-FTTH

Fig. 9. Contributing peers in Europe by ISP.

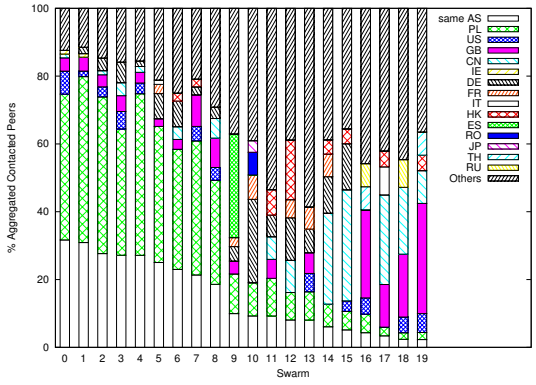


Fig. 10. External peer spatial distribution by swarm.

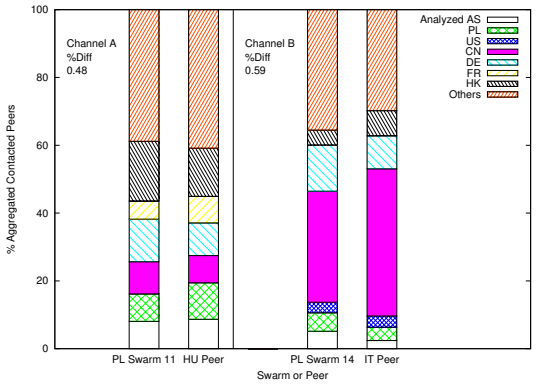


Fig. 11. External peer spatial distribution as viewed at different geographical points.

A. *Peer discovery process*

We start by showing geographical location of contacted peers during a given event for two different ISPs in Fig. 9. Top plot refers to a trace collected in Poland, while bottom plot refers to a trace collected in Italy. As it can be observed, the countries of peers interested in the event are very different. For instance lot of peers are found in Poland, Germany and U.K. in top plot, while bottom plot shows very few peers in Poland and U.K. Regardless subdivision of channels, from a global point of view seems that users from different countries will choose different channels.

To highlight this we focus on geographical distribution of external peers for different channels. The goal is to understand how different is the spatial distribution of peers in different swarms. Fig. 10 reports the breakdown of all external peers according to their origin country, identified by the Country Code. The set of countries has been selected among the most frequent ones. To easy readability, in this figure we sort swarms according to the fraction of peers that belong to the same Polish Autonomous System (labeled "same AS") the probe is located in.

Several considerations hold. First, the fraction of peers located in Poland is larger than 50% up to Swarm 8. Then, it suddenly drops to less than 20%. Note that the portion of peers in same AS with respect to the total Polish peers is very similar in all swarms. This reflects the market share of the ISP the probe was located in. Second, the fraction of peers found in other countries is variable. For example, Swarm 9 has a large fraction of users in Spain (and indeed this swarm corresponds to a football match involving a Spanish team), while Swarms from 16 to 19 have a clear predominance of GB users (and they all correspond to "Premier League" events). This clearly shows that there is a bias in the external peers contacted during each event which is naturally induced by the actual distribution of users; this, in its turn, is highly dominated by cultural and language traits.

The surprisingly large portion of peers in Poland may also be induced by SopCast, which could constraint the peer discovery process according to some "distance" metric. To check if SopCast actually enforces any peer geographical location bias, we compare the external peer distribution observed from two different countries. For the comparison, we select two swarms: in Swarm 11 we monitored peers in both Poland and Hungary; in Swarm 14 we monitored peers in Poland and Italy. Fig. 11 shows the distribution of peers as seen from the different probes. The two distributions are statistically very similar.

To quantify this similarity, we evaluate the *distance* between the two distributions by simply computing the difference among the number of peers in each country, and normalizing

it with respect to the total number of peers,

$$\sum_{CC} \left| \frac{|P(a) \in CC| - |P(b) \in CC|}{|P(a) \cup P(b)|} \right|.$$

The distance is equal to 0.48% and 0.59% for Swarm 11 and Swarm 14, respectively, meaning that there is a very large similarity among the peers spatial distributions. This allows us to conclude that SopCast peer discovery is not driven by any preference, but the natural distribution of users in the world. That is, the peer discovery follows a random process in which the probability of contacting (being contacted) by a peer is the same for all peers, but peers are non-uniformly distributed in space. The user preference for some events is the largest driver of peer availability. Recently, this bias has been observed for file sharing P2P system too [17].

B. Content retrieval

We are now interested to observe which peers contribute more on the download of content. Since peer discovery follows a random process, does also the content retrival follows a random choice?

To answer this question, we have compare the amount of traffic downloaded from internal peers from i) other peers in the same AS, ii) other peers in Poland, iii) other peers with large upload capacity. Results shows that there is a large preference to retreive content from peers that have a large upload capacity only. To illustrate this phenomenon, we focus on the amount of content provided by peers belonging to two external ASes that are known to offer high upload capacity to customers. One of these networks is the AS that provides connectivity to Universities and Research centers in Poland, while the second one is an AS located in Russia in which customers are offered links of 100Mb/s capacity.

Figs. 12-13 depict the fraction of peers found in these networks and of traffic received from and transmitted by the internal peers. Although the fraction of contributing peers in these networks is marginal (no more than 2%) the amount of transmitted traffic to internal peers is important (upto 10% of total traffic received). This highlights that high bandwidth peers are fundamental to sustain the SopCast service. But are Polish high bandwidth peers preferred?

To answer this second question, note that in the swarms that are popular in Poland (see PL aggregated contacted peers percentage in Fig. 11), it is easy to find Polish high bandwidth contributing peers. This has favorable implications for the geographical traffic allocation, since it remains enclosed in the same country, where often traffic exchange is cheaper than the one imported from abroad. On the other hand when high capacity peers are scarce in Poland, the need to find high bandwidth peers enforces the download of video traffic from other countries (e.g., from Russia for the second subset of channels reported in Fig.13).

SopCast thus fails enclosing traffic geographically by just using as policy the selection based on upload capacity. Thus cultural and language traits, play an important role on the traffic allocation that favors only to some networks.

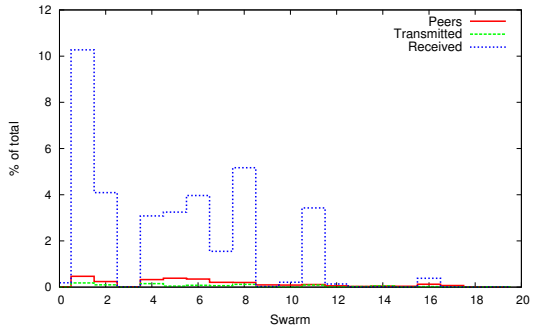


Fig. 12. PL High Capacity AS contribution.

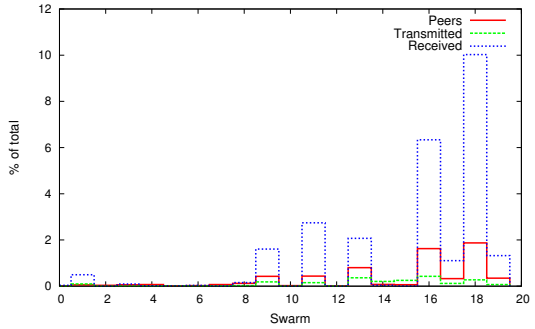


Fig. 13. RU1 High Capacity AS contribution.

VIII. Conclusions and Implication for Traffic Localization

In this paper we have presented measurements results obtained by monitoring backbone traffic for more than a year. We extensively analyzed P2P-TV traffic generated by customers of three different ISPs when using SopCast, PPLive or TVAnts applications.

Results considering users habits show that SopCast is the largely preferred system. Customers use it to watch sport events that are difficult or expensive to retrieve using traditional means. While the average traffic due to these applications is overall marginal, few tens of users generate 15% of traffic in the monitored PoP where more than 20.000 users are aggregated during popular events.

We then investigate SopCast swarm formation, focusing on peer discovery and data delivery processes. Our results indicate mainly two things. First, given the nature of P2P-TV, the distribution of peers, and, hence, traffic, has significant locality properties deriving from users' cultural and language traits. While this was already noticed in the case of file-sharing applications, we believe that for P2P-TV systems this phenomenon is even more visible. Second, data are mainly provided by high capacity peers and P2P-TV provision would not be feasible without their support, indeed ADSL connected peers can only contribute to about 1/5 of the required traffic.

This means that, even if some network aware peer selection mechanism is enforced to localize traffic to any subset of peers, its effectiveness might be limited by i) the already strong localization of peers enforced by users' preference and ii) the absence of sufficient capacity provided by the subset of peers.

For example, ISPs providing ADSL lines to their users that desire strong locality of traffic, might need to deploy a few high capacity peers, acting as super peers, inside their network.

REFERENCES

[1] D. Ciullo, M. Mellia, M. Meo, and E. Leonardi. Understanding P2P-TV systems through real measurements. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–6. IEEE, 2008.

[2] D. Ciullo, M.A. Garcia, A. Horvath, E. Leonardi, M. Mellia, D. Rossi, M. Telek, and P. Veglia. Network Awareness of P2P Live Streaming Applications: A Measurement Study. *Multimedia, IEEE Transactions on*, 12(1):54–63, 2009.

[3] B. Li, S. Xie, Y. Qu, G.Y. Keung, C. Lin, J. Liu, and X. Zhang. Inside the new coolstreaming: Principles, measurements and performance implications. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages 1031–1039. IEEE, 2008.

[4] C. Wu, B. Li, and S. Zhao. Multi-channel live P2P streaming: Refocusing on servers. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages 1355–1363. IEEE, 2008.

[5] X. Hei, C. Liang, J. Liang, Y. Liu, and K.W. Ross. A measurement study of a large-scale P2P IPTV system. *Multimedia, IEEE Transactions on*, 9(8):1672–1687, 2007.

[6] F. Wang, J. Liu, and Y. Xiong. Stable peers: Existence, importance, and application in peer-to-peer live video streaming. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages 1364–1372. IEEE, 2008.

[7] L. Vu, I. Gupta, J. Liang, and K. Nahrstedt. Measurement of a large-scale overlay for multimedia streaming. In *Proceedings of the 16th international symposium on High performance distributed computing*, pages 241–242. ACM, 2007.

[8] X. Hei, Y. Liu, and K.W. Ross. Inferring network-wide quality in P2P live streaming systems. *Selected Areas in Communications, IEEE Journal on*, 25(9):1640–1654, 2007.

[9] S. Agarwal, J.P. Singh, A. Mavlankar, P. Baccichet, and B. Girod. Performance and quality-of-service analysis of a live P2P video multicast session on the Internet. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on*, pages 11–19. IEEE, 2008.

[10] Y. Lu, B. Fallica, F.A. Kuipers, R.E. Kooij, and P.V. Mieghem. Assessing the quality of experience of sopcast. *International Journal of Internet Protocol Technology*, 4(1):11–23, 2009.

[11] T. Silverston and O. Fourmaux. Measuring p2p iptv systems. In *Proc. of ACM NOSSDAV*, 2007.

[12] T. Silverston, O. Fourmaux, A. Botta, A. Dainotti, A. Pescapé, G. Ventre, and K. Salamatian. Traffic analysis of peer-to-peer IPTV communities. *Computer Networks*, 53(4):470–484, 2009.

[13] E. Leonardi, M. Mellia, A. Horvath, L. Muscariello, S. Niccolini, and D. Rossi. Building a cooperative P2P-TV application over a wise network: the approach of the European FP-7 strep NAPA-WINE. *Communications Magazine, IEEE*, 46(4):20–22, 2008.

[14] M. Mellia, R. Lo Cigno, and F. Neri. Measuring IP and TCP behavior on edge nodes with Tstat. *Computer Networks*, 47(1):1–21, 2005.

[15] N. Cascarano, F. Risso, A. Este, F. Gringoli, L. Salgarelli, A. Finamore, and M. Mellia. Comparing P2PTV Traffic Classifiers. In *Communications (ICC), 2010 IEEE International Conference on*, pages 1–6. IEEE, 2010.

[16] http://www.maxmind.com. Geoip Database.

[17] R. Cuevas, N. Laoutaris, X. Yang, G. Siganos, and P. Rodriguez. Deep diving into bittorrent locality. In *Proceedings of the 5th international student workshop on Emerging networking experiments and technologies*, pages 7–8. ACM, 2009.