P2PTV Multi-channel Peers Analysis

Marwan Ghanem, Olivier Fourmaux
Sorbonne Universités, UPMC Univ Paris 06
CNRS, LIP6 UMR 7606
75005 Paris, France
Email: olivier.fourmaux@upmc.fr

Fabien Tarissan
Université Paris-Saclay
ISP, ENS Cachan, CNRS
94235 Cachan, France

Takumi Miyoshi
College of Systems Eng. and Sc.
Shibaura Institute of Technology
Saitama, 337-8570 Japan

***Abstract*—After being the support of the data and voice convergence, the Internet has become one of the main video providers such as TV-stream. As an alternative to limited or expensive technologies, P2PTV has turned out to be a promising support for such applications. This infrastructure strongly relies on the overlay composed by the peers that consume and diffuse video contents at the same time. Understanding the dynamical properties of this overlay, and in particular how the users switch from one overlay to another, appears to be a key aspect if one wants to improve the quality of P2PTV.**

In this paper, we investigate the question of relying on non-invasive measurement techniques to track the presence of users on several channels of P2PTV. Using two datasets obtained by using network measurement on P2PTV infrastructure, we show that such approach contains sufficient information to track the presence of users on several channels. Besides, exploiting the view provided by sliding time windows, we are able to refine the analysis and track users that switch from one channel to another, leading to the detection of super-peers and providing explanations of the different roles they can play in the infrastructure. In addition, by comparing the results obtained on the two datasets, we show how such analyses can shed some light on the evolution of the infrastructure policy.

I. Introduction

After being the support of the data and voice convergence, the Internet has become one of the main video providers (live TV or video on demand). Those multimedia services were previously confined to the video broadcasting infrastructures (terrestrial, satellite or hybrid fibre/coax). The transmission of broadcasting quality TV streams in High Definition (or soon in Ultra High Definition 4K/8K) requires the use of huge amount of communications networks resources. The development of dedicated technologies to distribute these contents is either local and limited to a residential operator (IPTV), or global but complex and expensive (CDN).

The alternative to these limited or expensive technologies could be partly or completely based on P2P. In this context, peers communicate via virtual mesh networks (overlays) that connect them. The dynamic topology of these overlays depends on many parameters: location of resources, network status, internal mechanisms of peers, as well as the distributed content and the behaviour of the peers directly involved as consumers of content.

In the case of TV streams, specific constraints require significant adaptation of P2P (specifically related to real-time aspects). A new application class realizes this kind of service: P2PTV. For these applications, the content consists of audio/video streams to distribute in real-time to a large number of receivers. The large number of streams and their intrinsic real-time characteristics generate timing constraints which are difficult to guarantee in the considered dynamic environment. Strict compliance with these constraints impacts directly on the peer's quality of experience and thus on his behaviour, which in turn impacts the overlay.

P2PTV applications broadcast hundreds of channels, each carrying a live audio/video content to thousands of peers. Each channel corresponds to an overlay integrating peers wishing to receive its contents, and these peers can switch channels at any time (usually depending on the contents) adding an extra dynamic factor.

This is this last dynamical aspect we intend to address in the present paper. Although several works have been proposed to measure and analyse the activity on P2PTV infrastructure, tracking the presence of peers active on different channels remains challenging. Here we show that we can rely on non-invasive measurement techniques such as Wireshark to track peers switching from one channel to another. To do so, we rely on 2 datasets obtained by measurements campaigns that coordinate several points of measure on a well known P2PTV infrastructure; we show that although the views obtained by such a measurement approach are partial, they are sufficient to detect multi-channel peers and highlight particularities in their behaviour, thus leading the way to a more in-depth investigations on the subject.

The remainder of the paper is organised as follow: we start by presenting existing works related to the analysis of P2PTV applications (Section II) before presenting the dataset used in the present paper (Section III). Then we turn to the study of the information contained in the dataset in order to analyse the behaviour of multi-channel peers. We start by exploiting the dataset by aggregating all the information (Section IV) before refining our analysis using narrowed views provided by sliding time windows (Section V). Then we show that comparing the two datasets gives insight on how diffusio through P2PTV has evolved (Section VI). Finally we conclude the paper by presenting the perspectives opened by the present study (Section VII).

II. Related Works

Several studies and experiments have been done to analyse P2PTV applications [1], [2], [3], [4], [5], [6], [7]. Rossi *et al.*

proposed a framework for comparing P2P applications [6] in which they define a set of observable features related to the protocols used by the applications. They highlighted the main similarities and differences between several P2P applications. In particular, they provided the key elements that open the way to passive analyses one can use when the applications are proprietary and no internal access is provided. Spoto *et al.* presented an investigation of PPLive using both active and passive measurements [3]. Using a crawler, they were able to classify the traffic into three classes as well as to show that only 15% of peers could be considered as active peers, revealing the potentials and limits of PPLive active measurement strategies.

Other works have been done on a more quantitative perspective [8], [9], [10], [11]. Hei *et al.* proposed for instance a large scale measurement study of P2PTV, using a PPLive dedicated crawler [8]. By collecting a huge amount of data in different scenario, they have shown that P2P IPTV peers have the same behaviour as the one of TV peers. They also demonstrated the existence of a small set of super-peers that highly contribute to the video uploading. Similarly, using a crawler, Jia *et al.* tried to characterize PPStream [9]. They were able to find certain characteristics such as geographical clustering, arrival/departure patterns and playback quality.

Magharei *et al.* proposed a study on the structure of networks that most P2PTV applications used. They examine key issues with such structures and how bottlenecks can appear [12].

By passively studying the traffic in P2PTV infrastructures, Silverston *et al.* were able to compare different applications pointing out their similarities and differences [13]. Looking more deeply into the traffic, they discovered that signalling traffic tends to have a large inter-packet time while video traffic tends to have a smaller one. They also looked into peer behaviour, revealing that the vast majority of peers tend to receive data more than they send, pointing out potential reciprocity issues.

There are also few studies more specific to the peer behaviour and the multi-channel observations [14], [15], [16], [17]. Wang *et al.* analysed the traffic that is characteristic to peers switching from one channel to another [16]. Using the most popular P2PTV applications such as PPLive or SOPCast, they monitored a channel for a given period and then suddenly changed to another one. They revealed that switching has a huge impact on the network efficiency as it increases the overload and adds a significant overhead. Finally, Mitzutani *et al.* were able to detect video servers as well as to find new characteristics of PPTV by monitoring multi-channel PPTV traffic [17].

III. DataSet

In this section, we present the datasets that were used during this work. It consists in two distinct measurement campaigns conducted on PPTV at a different time. The key aspect of those campaigns is that they coordinated traffic measurements from different points of measure. Concretely, the measurements were conducted on several PCs each running the application on a different channel thus, from the application's point of view, they acted as a regular peer. Every PC had an Internet connection provided by FLET'S HIKARI NEXT, 100 Mbps optical access service via Plata HIKARI Mate as an ISP in Japan. For capturing and monitoring traffic, Wireshark [18], a well-known packet sniffer, was running on every measurement PC during the campaigns. Therefore we have the totality of the traffic that has been sent and received by our machines.

Table I
PROPERTIES OF THE DATASET

Property	**2013**	**2015**
Duration	14 hours	7 days
Number of channels	12	10
Number of peers	100 809	289 710
Maximum number of peers per channel	21 518	96 258
Average payload size	504 B	408 B
Total payload size	193 GB	601 GB

The first dataset was extracted from a 14 hour long traffic measured on December 2013 using 12 points of measure, while the second was extracted from a 7 day long traffic measured on July 2015 using 10 points of measure. We shall refer to those datasets later on as 2013 and 2015 respectively.

Table I presents the global properties of both datasets. We can particularly notice the huge amount of data exchanged (193 GB and 601 GB for 2013 and 2015 respectively). It is also worth noticing that, although 2015 dataset is way longer (12 times longer than 2013 dataset), it exhibits a less dense traffic than expected (the total payload size is for instance only 5 times higher), which is partly due to the lower number of channels.

As mentioned in many previous works [19], traffic generated by such applications can be shared out into two categories. Control (or signalling) traffic which could be either a heart beat signal, a peer's list exchange or buffer maps in form of a bit vector, representing the data a peer has available or missing from its video buffer. The second kind of traffic is data (or video) traffic which is transferred in the form of data chunks.

Most P2PTV applications were initially designed as a P2P mesh-based architecture [12] including PPlive and PPStream [9]. Nowadays, most applications use hybrid P2P infrastructures with super-peers to guarantee that the viewers receive a better quality. Each transmitted channel by a P2PTV application would have its own P2P mesh-based network, which contains two types of peers. The first are the super-peers that are servers, are active the whole time and appear in more than one P2P mesh in the same slot of time with the goal of maintaining the infrastructure of such systems. The second type is the regular peers that might appear in more than one channel due to switching behaviours [16].

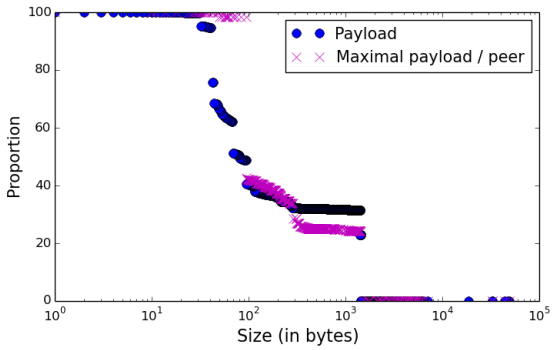


Figure 1. Inverse CDF of payload size and maximal payload size/peer

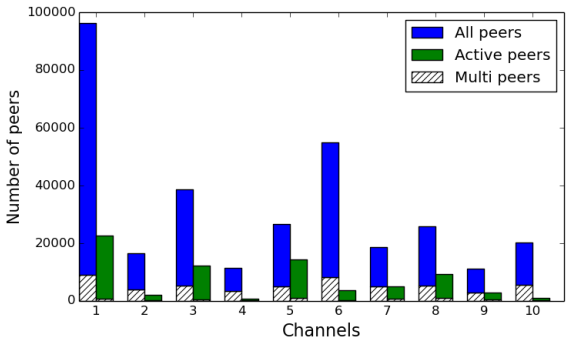


Figure 2. Distribution of the peers over the channels.

IV. Global analysis

In this section, we exploit all the information contained in the collected data in order to detect multi-channel peers. We start by presenting global properties of the dataset (IV-A) that complete those provided in Section III. Then we focus on the topics of this paper, namely the presence of multi-channel peers (Section IV-B). Note that in the rest of the section, if not mentioned otherwise, we will present the results on the 2015 dataset as it is the more recent and the larger one.

A. Tracking exchanges of video content

As mentioned in Section III, since we monitored the entire traffic using Wireshark, it is necessary to distinguish exchanges depicting the activity of a peer watching a TV program from the traffic dedicated to controlling the P2P infrastructure. To do so, it is reasonable to assume that a peer actively watching a TV program will trigger exchanges of video content, thus leading to a stream of packets with a significant size.

Figure 1 shows the inverse cumulative distribution of the payload size (plain circles). One can clearly observe two regions. The first region involves packets smaller than 1000 bytes (68%) while a second region involves packets larger than 1000 bytes (32%). Obviously, the first one is related to control traffic while the second one can be categorized as video exchanges. In order to simplify the analysis, we will further make the assumption that any packet whose size is less than 1000 bytes is not a video content. Furthermore, a peer involved in at least one traffic containing a video will now be referred to as an *active peer*.

In order to get a better image of the peers behaviour, we also display on Figure 1 the inverse cumulative distribution of the maximal payload size a peer has exchanged (cross dots). In other words, a (x, y) dot in this plot indicates that $y\%$ of the peers have exchanged packets with x bytes at most. This plot reveals that only 25% of the peers are actively involved in video traffic

B. Presence multi-channel peers

Turning now to the analysis of multi-channel peers, we start by studying the proportion of peers identified in several channels. It is worth remembering that the dataset consists of partial and independent measurements of the 10 channels (Section III). Thus, it is absolutely not guaranteed to detect such a behaviour.

Figure 2 presents the distribution of the peers over the 10 channels. For each channel, we show the number of all peers detected (left bar in blue) and the number of active peers (right bar in green) as defined above. In addition, we show for each of these quantities the fraction of the peers that are also detected in at least one other channel (bottom part of the bars with hatched lines). We will therefore refer to those peers as *multi-channel peers* (or multi peers for short).

The chart shows that the majority of the peers are concentrated in three channels. More importantly, it answers the first question raised in this paper which confirms that the measurement approach on which we rely on enables us to detect peers that appear in several channels. It turns out that 8% of the peers are multi-channel peers. Moreover, one can notice that this statement still stands even if we focus only on peers that exchange video contents, although the ratio then drops to 0.8% of the total number of peers. However, the ratio of multi-channel peers actively watching a TV program still involve 3% of the active peers, thus revealing that a non negligible fraction of peers exchanging video content are involved in several channels. Note that this percentage could be overestimated if distinct active peers are recorded with a single identifier, which would be the case if they are behind a NAT. Although we did not investigate deeply this question in a systematic manner, we manually looked in detail the most susceptible peers of the dataset. Our preliminary results show that this is not the case and that those IPs detected on several channels depict a unique peer.

V. Exploiting sliding time windows

The results presented in the previous section are interesting as they highlight the presence of multi-channel peers but aggregating all the information contained in the dataset prevents further refinement regarding the real behaviour of the peers. In particular, it does not allow to distinguish a peer that switches between different channels (referred to further as a *switching peer*) from a peer that stops watching TV programs and starts watching another one way later on.

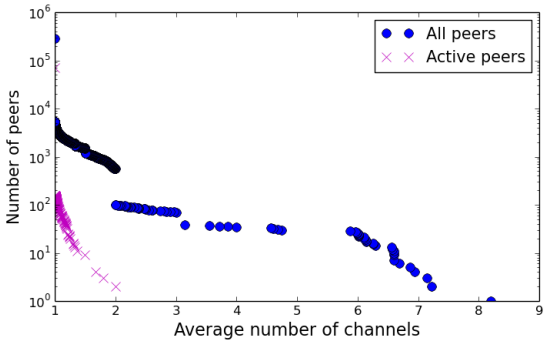


Figure 3. Inverse CDF of the average multi-channel presence.

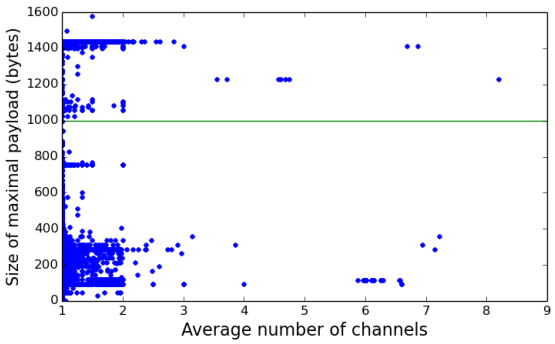


Figure 4. Average multi-channel presence *v.s.* maximal payload size

To overcome this issue we propose in the present section to rely on the view provided by sliding time windows. More precisely, we sliced the whole dataset into non-overlapping windows of similar size and studied whether it enables an in-depth analyses of multi-channel peers. As one can expect, the size of the window becomes a key parameter in this approach. Since we focus on tracking the presence of switching peers, the size has to be short enough to discard peers that disconnect but it also has to be long enough to be able to detect the presence of the peers in several channels. Therefore we decided to use a 1 minute size window.

In the following sections, we will investigate how this approach enables us to distinguish between different types of peers (Section V-A) and different types of super-peers (Section V-B).

A. Different peer behaviour

By relying on short-time windows, we are now able to detect peers present *simultaneously* on different channels. In particular, we can determine the number of peers present in different channels for each slot of 1 minute. Besides, when such a peer is detected, one can track how many channels it is involved in.

Figure 3 presents the inverse cumulative distribution of the average number of channels on which a peer is simultaneously present; for all the peers (plain circles) and for active peers (cross dots). In both cases, there is a large amount of peers involved in 1 channel only, then the value decreases smoothly between 1 and 2. However, when it comes to values higher than 2, active peers (cross dots) are no longer present. While in case of all peers (plain circles), a few peers are close to 4 or 5 channels in average. Finally, it is remarkable that some peers appear in almost all the channel (especially when computing the average value) when using such a short-time window. This indicates an unusual behaviour and it is reasonable to consider such peers as *super-peers*, *i.e.* usually servers active the whole time whose purpose is to maintain the efficiency of the infrastructure (see Section III).

B. Different super-peers behaviour

The detection of *video injectors* raises the question of the role of super-peers present on several channels in the infrastructure. In particular, do they all participate actively to the diffusion of video content or do they also regulate and control the traffic over the P2P infrastructure?

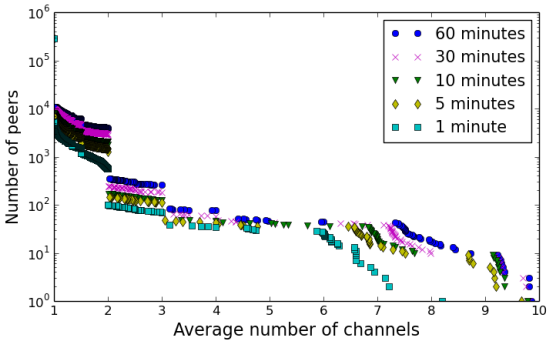


Figure 5. Inverse CDF of the average multi-channel presence for different time windows.

In order to have a better understanding of this question, we compare in Figure 4 the average number of channels in which a peer is simultaneously present to the maximal size of a payload it sent/received. Each dot standing for a different peer, one clearly finds the super-peers detected previously but we can now refine their role: obviously, most of them only support control traffic since all the packets they exchange have a very low size. Thus, these super-peers are present only for administration and surveillance purposes, while the *video injectors* are clearly present in the top left part of the figure mixed with regular users.

VI. Extending the analyses

In order to strengthen the results presented so far, we present in this section different analyses that allow to sustain our former claims. In particular, we show that the size of the time window has little impact on the former conclusions (Section VI-A). We also compare the results obtained on the 2015 dataset to the ones obtained on the 2013 dataset (Section VI-B). This allows to give insights on how diffusion through P2PTV has evolved.

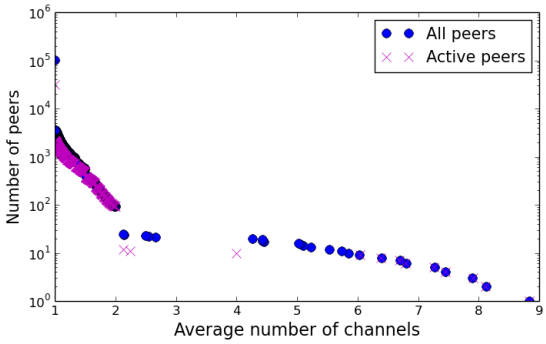


Figure 6. Inverse CDF of the average multi-channel presence (2013 dataset).

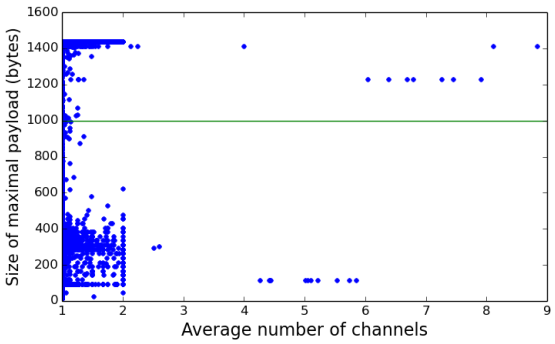


Figure 7. Average multi-channel presence *v.s.* maximal payload (2013 dataset).

A. Impact of the size of the window

All the results presented in Section V are strongly related to our choice of using a 1 minute time window. Although we claim that this choice is reasonable regarding our main objective (tracking multi-channel peer), one might wonder whether another value would have altered our conclusions. We therefore present in Figure 5 a figure similar to Figure 3 but for different time windows (1, 5, 10, 30 and 60 minutes). The figure shows that the main effect of increasing the size of the time window is to increase the number multi-channel peers. This is completely expected since more times gives more opportunity to switch between channels. But if the size of the window impacts quantitatively the results, the overall observations remain qualitatively valid. In particular, all curves present a sharp breach around the value of 2 and another one after 3. It is worth noticing that increasing the size of the time window also reveals that some super-peers cover all of the ten channels. Thus, this confirms our assumption that they do not correspond to regular peers.

B. Comparisons of the datasets

We turn now to the comparison between the two datasets. Using the same exact method explained in Section V-A), Figure 6 presents the inverse cumulative distribution of the average number of channels on which a peer is simultaneously present. Unlike Figure 3, the breach between values 2 and 4 is more prominent. More interesting is that we can see that super-peers are also present on the active peer curve (cross dot). Thus, we can deduce that between 2013 and 2015 the roles of super-peers have changed. In 2013, a super-peer could have been responsible for injecting video contents as well as administrating the infrastructure. In contrast, in 2015 each of these two roles seems to be taken in charge by a specific super-peer.

Finally, Figure 7 compares the average number of channels to the maximal payload size it has sent/received. We compare the figure with the one obtained on the 2015 dataset. Like in dataset 2015, the majority of super peers involved in several channels support only control traffic. However, in the 2013 dataset several super peers also support video traffic, which is in sharp contrast with what can be observed in the 2015 dataset. This is another indication of the important modification that took place between the two measurement campaigns.

VII. Conclusion

In this paper, we used two datasets obtained by two measurement campaigns relying on respectively 10 and 12 points of measure associated to 10 and 12 channels on a P2PTV system. We investigated how much information we can retrieve on multi-channel peers and showed that although the obtained view is partial, such a non-invasive measurement approach yet enables to detect peers present in several channels. Indeed, we were able to detect that respectively 8% and 16% of the total number of peers were present on more than one channels during the measurement campaign. This number drops to 2% and 3% if we restrain to active peers, *i.e.* peers involved in video content traffic. In addition, conducting an analysis based on sliding time windows led to precisely track peers switching from one channel to another one as well as to identify super-peers and to qualify their role in the infrastructure.

These results are interesting to characterize the behaviour of peers and extrapolate the behaviour of users. This is important for content providers to adapt the video supply, and for network operators to optimize their infrastructure.

These results also allow to envision promising lines of research. Since the measurement approach is light and does not require to have privilege access to the application itself or its infrastructure, we intend to conduct several measurement campaign targeting both different P2PTV systems and different measurement points to detect more characteristics related to the applications or to the users behaviour.

Acknowledgements

This study was partly supported by JSPS KAKENHI Grant Number 26420370. The datasets were captured by Koki Mizutani (Dataset 2013) and Naoki Takeuchi (Dataset 2015).

References

[1] S. Agarwal, "A case study of large scale p2p video multicast," in *IP Multimedia Subsystem Architecture and Applications, 2007 International Conference on*, Dec 2007, pp. 1–5.

[2] E. Alessandria, M. Gallo, E. Leonardi, M. Mellia, and M. Meo, "P2p-tv systems under adverse network conditions: A measurement study," in *INFOCOM 2009, IEEE*, April 2009, pp. 100–108.

[3] S. Spoto, R. Gaeta, M. Grangetto, and M. Sereno, "Analysis of pplive through active and passive measurements," in *Parallel & Distributed*, 2009.

[4] S. Valenti, D. Rossi, M. Meo, M. Mellia, and P. Bermolen, "Accurate, fine-grained classification of p2p-tv applications by simply counting packets," in *Traffic Monitoring and Analysis*. Springer, 2009, pp. 84–92.

[5] S. Tang, Y. Lu, J. M. Hernández, F. Kuipers, and P. Van Mieghem, "Topology dynamics in a p2ptv network," in *NETWORKING 2009*. Springer, 2009, pp. 326–337.

[6] D. Rossi, E. Sottile, and P. Veglia, "Black-box analysis of internet p2p applications," *Peer-to-Peer Networking and Applications*, vol. 4, pp. 146–164, June 2011.

[7] P. Bermolen, M. Mellia, M. Meo, D. Rossi, and S. Valenti, "Abacus: Accurate behavioral classification of p2p-tv traffic," *Computer Networks*, vol. 55, no. 6, pp. 1394 – 1411, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128610003713

[8] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross, "A measurement study of a large-scale p2p iptv system," *IEEE Transactions on Multimedia*, vol. 9, no. 8, pp. 1672–1687, December 2007.

[9] J. Jia, C. Li, and C. Chen, "Characterizing ppstream across internet," in *Proceedings of the 2007 IFIP International Conference on Network and Parallel Computing Workshops*, ser. NPC '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 413–418. [Online]. Available: http://dl.acm.org/citation.cfm?id=1306873.1307128

[10] T. Hoßfeld and K. Leibnitz, "A qualitative measurement survey of popular internet-based iptv systems," in *Communications and Electronics, 2008. ICCE 2008. Second International Conference on*. IEEE, 2008, pp. 156–161.

[11] D. Wu, Y. Liu, and K. Ross, "Queuing network models for multi-channel p2p live streaming systems," in *INFOCOM 2009, IEEE*, April 2009, pp. 73–81.

[12] N. Magharei and R. Rejaie, "Understanding mesh-based peer-to-peer streaming," in *Proceedings of the 2006 International Workshop on Network and Operating Systems Support for Digital Audio and Video*, ser. NOSSDAV '06. New York, NY, USA: ACM, 2006, pp. 10:1–10:6. [Online]. Available: http://doi.acm.org/10.1145/1378191.1378204

[13] T. Silverston, O. Fourmaux, A. Botta, A. Dainotti, A. Pescapé, G. Ventre, and K. Salamatian, "Traffic analysis of peer-to-peer iptv communities," *The International Journal of Computer and Telecommunications Networking*, vol. 53, no. 4, pp. 470–484, March 2009.

[14] M. Cha, P. Rodriguez, J. Crowcroft, S. Moon, and X. Amatriain, "Watching television over an ip network," in *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '08. New York, NY, USA: ACM, 2008, pp. 71–84. [Online]. Available: http://doi.acm.org/10.1145/1452520.1452529

[15] J. Mendes, P. Salvador, and A. Nogueira, "P2p-tv service and user characterization," in *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, June 2010, pp. 2612–2620.

[16] M. Wang, O. Fourmaux, Y. Nakamura, and T. Miyoshi, "Network impact of p2p-tv zapping." IEEE/ACIS-SNDP, July 2013.

[17] K. Mizutani, T. Miyoshi, and O. Fourmaux, *Traffic Analysis in Concurrent Multi-Channel Viewing on P2PTV*. Springer Berlin Heidelberg, 2015.

[18] "Wireshark." [Online]. Available: http://www.wireshark.org/

[19] X. Hei, Y. Liu, and K. W. Ross, "Iptv over p2p streaming networks: the mesh-pull approach," *Communications Magazine, IEEE*, vol. 46, no. 2, pp. 86–92, February 2008.