Modeling and Analyzing the Video Game Live-Streaming Community

Gustavo Nascimento*, Manoel Ribeiro*, Loïc Cerf*, Natália Cesário*, Mehdi Kaytoue†,
Chedy Raïssi‡, Thiago Vasconcelos* and Wagner Meira Jr.*
* *Universidade Federal de Minas Gerais, Departamento de Ciência da Computação, Belo Horizonte, Brazil,*
Email: {gunasper, manoelribeiro, lcerf, natalia, thiagorova, meira}@dcc.ufmg.br
† *Université de Lyon, CNRS, INSA-Lyon, LIRIS, UMR5205, F-69621, France,*
Email: mehdi.kaytoue@insa-lyon.fr
‡ *INRIA Nancy Grand Est, 54500, Vandœuvre-lès-Nancy, France,*
Email: chedy.raissi@inria.fr

***Abstract*—In parallel to the exponential growth of the gaming industry, video game live-streaming is rising as a major form of online entertainment. Gathering a heterogeneous community, the popularity of this new media led to the creation of web services just for streaming video games, such as Twitch.tv.**

In this paper, we propose a model to characterize how streamers and spectators behave, based on their possible actions in Twitch and, using it, we perform a case study on the Starcraft II streamers and spectators. In the case study we analyze a large amount of data collected in Twitch.tv's chat in order to better understand how streamers behave, and how this new form of online enteirtainment is different from previous ones.

Based on this analysis, we were able to better understand channel switching, channel surfing, and to create a model for predicting the number of chat messages based on the number of spectators. We were also able to describe behavioral patterns, such as the mass evasion of spectators before the end of a streaming section in a channel.

***Keywords*-video game; starcraft; streaming; twitch.tv;**

I. Introduction

Over the last decade, the annual turnovers generated by the electronic entertainment industry went beyond those of both cinema and music industries, making video game production a highly profitable business. A recent example is the announcement of a budget of US$ 500 millions for the video game Destiny (Blizzard/Activision) to be released in September 2014[1] eclipsing the previous record, Grand Theft Auto V (Rockstar Games), whose budget was US$ 265 millions and had an realease date earning of US$ 800 millions[2]. As a matter of fact, video games are challenging Hollywood blockbusters.

In parallel with the game industry exponential growth, a very related yet totally different form of online entertainment is also expanding at a very fast pace. Watching video-game live-streams is becoming an increasingly popular way of entertainment. From casual players recording themselves playing indie games to highly competitive e-sports international championships, the content produced has attracted a huge and heterogeneous community all over the world [1]. This may sound very peculiar at first glance as video games are primarily designed for the players, but so were designed physical sports with millions of weekly spectators, such as football or basketball.

This scenario has led to the creation of live-stream web services just for video games, like Twitch.tv. Started in 2011 from a Justin.tv fork, Twitch has grown quickly and met a huge success [2], [3]. In 2014, Twitch counts more than 45 million unique viewers per month gathered around for broadcasting, watching, and chatting from everywhere they play. The web service holds a very diverse set of sub-communities such as the *e-sports*, focused on highly skilled players and huge championships [4], the *speed running*, where the main objective is to finish a game as quickly as possible (often, beat a world record), and the *let's play*, characterized by charismatic streamers that interact a lot with his or her spectators [5]. Twitch is also used to stream (and promote) beta games, plays the role of targeted advertising of all kinds of games, and was included in several games consoles, such as PlayStation 4 and X-Box One systems. In February 2014, Twitch was 4^{th} in Peak US Internet Traffic, being responsible for 1.8% of the traffic, just behind Netflix, Google and Apple [6]. Understanding who produces and consumes this important volume of video content is as such a major interest. In summary, there is as clear a trend towards streaming ones games, as there is for one tweeting about his life.

In 2010, Kaytoue et al. made a first characterization of the e-sport live-streaming sub-community within Twitch [1]. In this paper we propose a model to analyze how the streamers and spectators behave and further characterizing the live-streaming community through a case study. The model consists of mapping the actions that spectators and streamers can perform as transition graphs, enabling us to analyze qualitatively and quantitatively the video game

[1]http://en.wikipedia.org/wiki/Destiny_%28video_game%29
[2]http://en.wikipedia.org/wiki/Grand_Theft_Auto_V

978-1-4799-6953-1/14 $31.00 © 2014 IEEE
DOI 10.1109/LAWeb.2014.9

live-streaming community. Using the proposed model, we are able to answer three fundamental questions about the community: (i) How do streamers and spectators behave? (ii) Are there patterns in those behaviors? and (iii) How is the content in the video game live-streaming community different from other kinds of online entertainment?

In our case study, we analyze the Starcraft II streamers and spectators, clearly part of the e-sports sub-community. Already studied in other papers [1], [7] and books [4], Electronic Sports started to really develop in the 90's with licenses like Doom and Counter Strike. Starcraft Broodwar (Blizzard Entertainment) was a huge success in South Korea: competitions were even casted on TV on prime time. Its successor, Starcraft II, met similar success, having its own world-wide player ranking system (ELO) and annual world cup competition series (WCS) with a US$1,6 million prize pool for the year 2014. Electronic sports would not be so developed if they were not supported by strong and active communities around the world. Reaching such communities is not possible though classic media. On the other hand, the usage of Social TV, or live streaming, works as a catalyst, that is, a mechanism to meet, discuss, and share the passion about e-sport.

Analyzing the data related to this fascinating environment, we discovered a lot about the behavior of both streamers and spectators. For instance, we were able to: (i) understand the meaning of switching from one streamer channel to another, (ii) describe the mass evasion of spectators that happens in the last minutes before and after the end of a streaming session, (iii) identify channel surfing as a behavioral pattern, and (iv) create a model to predict the amount of chat based on the number of spectators logged into a channel. We believe that these results are of major interest for the whole gaming community, and a large step into fully understanding this new forms of entertainment that are e-sports and video game live-streaming.

The rest of this paper is organized as follows. The second section discusses related works. In the third section we present some background concepts to the video game live-streaming context. In the fourth section we present our model for the video-game live-streaming community. In the fifth section we describe a case study with the Starcraft II streamers' channels, trying to answer the fundamental questions mentioned before. Finally, in the sixth and last section we conclude the paper.

II. Related Work

A first analysis of live streaming workloads on the internet can be found in [8]. We are more specifically interested here in the so called social TVs, which combine communication and social interactions in a TV framework [9], [10]. As the video game live-streaming is a kind of social TV, allowing interaction of the users via chat, the subjects are rather related. In [11] the social television model is described and analyzed, providing a framework to understand the current situation of social TV and identifying future developments. Similarly, [12] describes empirical results of computer-mediated groups using social TV, characterizes its model and suggests features to future social TV prototypes. Despite being closer to the traditional television content, both works provide valuable information about how the spectators interact in social TV, which can be applied to the video game live-streaming scenario.

Video games were also studied in the literature. In fact, not only the games themselves but how the players play those games, which was the most frequent issue. Despite differences from our work in this paper, [13] discusses how players learn to be grandmasters in games like Starcraft II and chess. They observed the entire training process and interviewed actual players in order to grasp the main aspects that lead a novice player to become a grandmaster. Whereas there are plausible cognitive markers of expertise that can be identified from the games logs (recording all players actions) [14], we may link the work of [13] and of [7] to find out that there are many novice players who learn by watching live stream games from grandmasters.

Finally, there are some works about both e-sports and video game live-streaming. [4] discusses multiple facets of the e-sports scene: pro-gaming, its highly paid players, play-by-play broadcasts, and mass audience, it also describes the whole e-sports environment of leagues, teams, organizers, sponsors and fans. In [1] we made a first characterization (based on collected data) on the e-sport community. In [7] the nature of the Starcraft II spectator (and from spectatorship itself) is studied; the paper proposes personas that represent the different reasons for people watching gamers playing, and the relevance of such habit to the larger video game live-streaming context. In [5], a broader view of the video game live-streaming context is presented. The work discusses the web services for live-streaming and describes the major sub-communities.

Our article has significant differences from the aforementioned related work. We perform a significant data analysis, in contrast with [4], [5], [7], which are quite theoretical, while proposing a model that captures the semantics of streamers and spectators interaction, going beyond data-driven analysis [1]. Further, we focus on the e-sports sub-community as [1], [4] but the model we propose is generic and may be used to analyze any e-sports sub-community. Our paper is also different from those about social TVs and from the gaming content, because the fusion of those two elements creates a completely different scenario which is the video game live-streaming.

III. Video game live-streaming

Before introducing our characterization model, we need to present, in this section, some background on video game live-streaming.

Twitch.tv is a platform that provides channels where users stream themselves playing and other users may watch them. The website also allows users to chat in real time in a given channel as well as provides an API that can be used to gather data about the streams. Spectators may choose channels using a search engine, browse the featured channels, or view channels by game or broadcaster. Twitch.tv makes money by placing advertisements over streamed content and by featuring sponsored channels, but it also allows streamers to monetize their streams. Spectators may, for example, subscribe to the channel of a given streamer to support his or her work.

The streamer is the user who streams live content on his or her channel—a web page in Twitch. The streamer may perform three actions in Twitch: (i) start a stream, (ii) prepare to end a stream, (iii) end a stream. Notice that preparing to end the stream is when the streamer signals that he or she will end the stream. In those moments the streamer bids farewell to the spectators and makes acknowledgements. We decided to highlight this action because we observed that it is usual that a significant number of spectators leave the channel when the streamer prepares to end the stream. The majority of the streams are generated by single persons, but teams or even championships are often streamers. An important observation is that, in Twitch.tv, each streamer has a channel where spectators may join regardless of the streaming being active.

The chat is an important part of the video game live-streaming experience because it enables both the spectator/spectator and the streamer/spectator interactions. The conversation subjects in the chat vary significantly, being very dependent on its associated sub-community, ranging from comments about the streamer performance to strategies about the game being played [5].

We define session as the time interval between the moment when the streamer starts to stream and the moment he or she ends the stream. Many streamers maintain a schedule of his or her sessions so that spectators know when his or her channel is online.

The spectator is the user who watches streams and may perform 4 actions in Twitch: (i) join a channel, (ii) leave a channel, (iii) send messages in the chat of a channel, and (iv) switch from one channel to another. The last action may be seen as a joint action that consists of leaving the channel of a streamer X and joining the channel of a different streamer Y. In [7] several stereotypes of video game live-streaming spectators were described. Examples of those stereotypes are people who watch professional players to improve their own gaming techniques and people who enjoy watching other people playing more than actually playing the game. Distinct spectators watch gaming live streams for distinct reasons.

IV. Our model

In this section we propose a model to characterize how streamers and spectators behave based on their actions in Twitch. In particular, we propose a transition graph for each role based on the actions described in Section III. The vertices in the transition graphs are states and the transitions are actions.

In our model, a streamer α is in one of three possible states: (i) streaming (ON), (ii) disconnecting (OD), that is, preparing to end the stream, and (iii) not streaming (OFF). We may then define the actions that α may perform as transitions between states:

Action	Src.	Dest.
Start a stream α	OFF	ON
Prepare to end a stream α ($\sim 3m$ before)	ON	OD
After ending a stream α ($\sim 15m$ after)	OD	OFF

For modeling a spectator S, we should take into account that he or she may have joined multiple channels, so there is a transition graph S_α for each pair streamer and spectator. S_α has two states: (i) connected (IN) and (ii) disconnected (OUT), depending on whether or not S has joined the channel. Next we present the actions of S and respective transitions:

Action	Src.	Dest.
Join channel	OUT	IN
Leave channel	IN	OUT
Chat	IN	IN

The transition graphs for both streamers and spectators are depicted in Figures 2 and 1, respectively.

Using this model, we may record the dynamics of Twitch.tv as a sequence of actions, where each action is a tuple $< G, s, a, d, t >$, where G is a transition graph (associated with either streamer or pair streamer-spectator), s is the source state, a is an action, d is the destination state, and t is a time-stamp.

For instance, we may easily represent when a spectator S switches from channel α to channel β through the transitions $< S_\alpha, IN, leavechannel, OUT, t_1 >$ and $< S_\beta, OUT, joinchannel, IN, t_2 >$. In this case, we may also define that a channel switch must happen within a time interval thr, which is easily formalized by the constraint $|t_1 - t_2| \leq thr$.

The model also allows us to analyze using data from both transition graphs, analyzing the states of the streamers involved when a spectator performs a set of actions. Using this, we may add another level of depth to the switch representation. We define an $A \rightarrow B$ switch as the set of transitions previously mentioned where the spectator leaves a channel where the streamer is in the A state, and joins a channel where the streamer is in the B state.

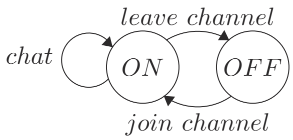

Figure 1: Spectator finite state machine.

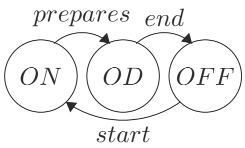

Figure 2: Streamer finite state machine.

We may also use this same idea to analyze the time period during which a spectator joins and leaves a streamer channel while considering the states the streamer was in both moments. We define a sojourn $A \approx B$ as the period of time between a spectator S joins α in state A, and leaves α in state B. Formally, given the sequence of transitions performed by S_α, where the ith transition is join channel, and the $i+1$th transition is leave channel, that is, $< S_\alpha, OUT, joinchannel, IN, t_i >$ and $< S_\alpha, IN, leavechannel, OUT, t_{i+1} >$, we state that $A \approx B = t_{i+1} - t_i$.

V. Starcraft II: a case study

Despite the fact that different games are associated with different workloads and behaviors on Twitch.tv, we believe that it is valuable to characterize and understand in more depth the community behind one specific game, since such understanding would help to grasp invariants that may help characterize other communities as well. We chose Starcraft II, a real-time strategy video game (RTS) with a large community, which has already been analyzed in the litterature [1], [7].

Starcraft II is the fastest selling real-time strategy game of all time.[3] It takes place in a science fiction environment. In that game, every player leads a so-called "race" (among three: Zergs, Protoss or Terrans), exploits resources located on the map (to create buildings, military units and get technological upgrades) and fight the opponents. The player must destroy the buildings of all the enemies to win the game. The victory requires mastering both sophisticated strategies (in an uncertain and real time environment) and a rapid low-level management of the units.

When choosing Starcraft II for our case study, we are actually taking two decisions: (i) to gather data from a single game, and (ii) choosing Starcraft II to be that game. The first decision reduce our sample space but diminishes the noise that gathering data from multiple games would implicate, due to the heterogeneity of Twitch sub-communities [7]. The second decision was taken because Starcraft II was already analyzed in other papers [1], [7], and because it is a game with a solid spectator community.

[3] http://www.eurogamer.net/articles/2010-09-01-starcraft-ii-sells-3-million-in-a-month

spectator	spectator single id.
channel	streamer single id.
date, time	the date and time the tuple was collected.
action	action performed by the user; the action can be joining a channel (INC), leaving a channel (OUT), or sending a message in the chat (CHAT).
chat	the message sent on the chat if the action is CHAT; empty for INC or OUT.

Table I: Elements in a tuple.

Collection period:	**02/10/13** to **17/02/14**
#spectators:	$1,460,740$
#channels:	136
#INCs:	$20,188,434$
#OUT:	$20,195,358$
#CHAT:	$13,878,122$
#tuples:	$54,247,484$
#sessions:	$4,944$
average session duration:	5,2h
median session duration	3,7h

Table II: General information about the dataset.

A. *Dataset*

Our analysis focuses on the Starcraft II players who reached, at some point, more than one thousand spectators, and the spectators watching those players. The REST API of Twitch.tv allowed us to identify those popular players. The dataset consists of the states of the streams (as given by the REST API) along 139 days and the logs of the chats associated with those streams. An IRC client, written in Python, collected the chat data. Table II summarizes our dataset. It is a set of $54,246,484$ tuples in the format $< user, channel, date, time, action, chat >$, as detailed in Table I.

Importantly, the dataset contains only registered users, i.e., users with an account on Twitch.tv. The remaining spectators cannot chat and no information on them is available but their total number on every channel.

B. *Pre-processing the data*

The rough data, for a given streamer, suffer from a few inconsistencies: some CHAT actions before the INC actions of the related users, some INC actions of users who are already connected (i.e., several INCs without OUTs in between), some OUT actions of users who are already disconnected (i.e., several OUTs without INCs in between) and users who appear connected but are inactive and never disconect.

The inconsistencies seem to occur because of the IRC's delay when delivering messages, once we consider the time when messages arrive and not when they're sent. On another hand, we also got some errors when our client is rebooting and some OUT actions are lost. The last scenario occurs 2 times a day and its duration is about 2 minutes.

The inconsistencies above affect less than 1% of the data. However, we decided to solve them, applying the following fixes (in the reported order):

1) whenever an apparently offline user chats, add an INC action just before (same time) its first CHAT action;
2) whenever there are several INC actions that occur in less than 1 minute without an OUT action in between, only keep the first one (notice that the fix at step 1 frequently creates this issue, once the real INC action uses the appears some seconds late);
3) whenever there are several INC actions that occur in more than 1 minute without an OUT action in between, add an OUT action after (same time) the last CHAT action. Notice that if she didn't write anything, nothing occurs;
4) whenever there are OUT actions without an INC action in between, or INC actions without an OUT action in between, remove the action.

C. The OD State

By definition, the OD state starts little before the disconnection of the streamer and ends little after that disconnection. However how "little" before and after? We decided to answer this question by analyzing the data. We assume that the OD state coincides with the spectators leaving the channel (hence the chat) at a higher rate.

For every session, we map the maximum number of users the session reached to 1 and proportionally compute a number between 0 and 1 for every other timestamp. After this per-session normalization, the numbers for all sessions (of all streamers) at the same time distance of the disconnection (considered as time 0) are considered and the median value is kept. Fig. 3 represents those median values around the disconnection (10 minutes before and 20 minutes after). About three minutes before the disconnection, the spectators start to leave the stream at a high pace. That value is therefore chosen as the start of the OD state. The leaving rate then decreases and the end of the OD state is not that clear from the data: some spectators keep on chatting on a channel hours after it ended. We chose 15 minutes as the end of the OD state. It corresponds to 30% of the maximum number of spectators who are still connected. The now formally defined OD interval is emphasized in Fig. 3.

That result justifies the existence of a new OD state in the transition graph of a streamer. The more complex transition graph supports a finer analysis of the behavior of the spectators: leaving a channel in the OD state is semantically different from leaving when it is either in the ON or the OFF state.

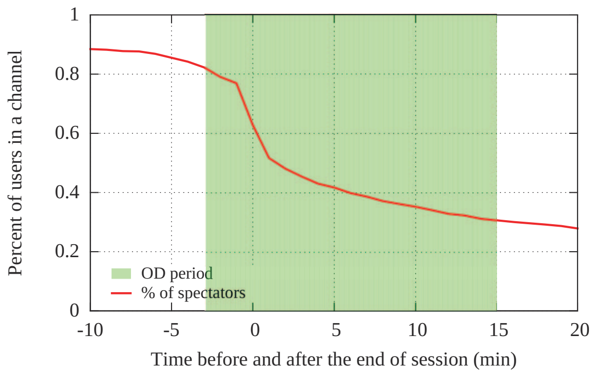

Figure 3: Spectators leaving a channel during the OD state.

D. Understanding the Streamers

The analysis in this section aims to better understand the behavior of the streamers and to see to what extent the content they produce differ from other kinds of online entertainment.

1) Session length: The lengths of the streaming sessions are an aspect of the content broadcast on Twitch.tv. It has already been studied in [1] but we focus here on the popular Starcraft II channels, whereas [1] studied the whole range of video games played by streamers of any level. therefore is worth doing, because it targets a different streamer population.

The average session length in the dataset lasts 5.2 hours. Its median is 3.7 hours (as stated in Table II). In [1] the session median was 1.58 hours. The discrepancy probably reflects the difference of player levels. The professional content, studied in this article, tends to last longer.

Fig. 4 shows the average length duration of each type of channel: individual player or not (teams or TV broadcasting competitions). Such type was manually inferred through a visit to every studied channel. Only 13% of the sessions last less then 2 hours. Most of the sessions last between 2 and 4 hours. The longest sessions (by far: 16 hours on average) come from TV channel broadcasting the World Championship Series[4].

It is interesting to notice how much watching video games differ from other popular kinds of online entertainment. As we have just seen, video game live-streaming produces long duration content in the same online world where YouTube's videos last 4 minutes and 12 seconds on average [15] and Twitter's messages are, at most, 140 characters long.

2) Streamer assiduity: Fig. 5 shows distributions of the average number of sessions per day. The maximal value that

[4]http://wcs.battle.net/sc2/en

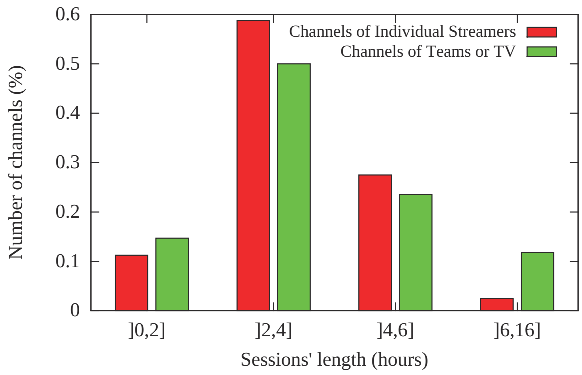

Figure 4: Distributions of sessions' length w.r.t. the type of channel.

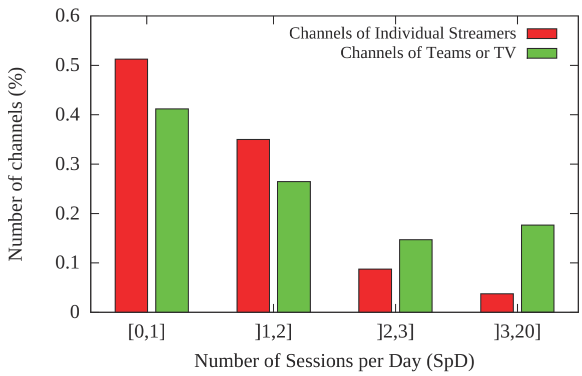

Figure 5: Distribution of sessions per day

is observed is an average of 19 sessions per day. Team and TV channels tend to broadcast more sessions than individual channels. That is expected since they stream more than one player. In fact, the higher the number of sessions per day, the more likely we are dealing with a team or a TV channel. Almost no individual players manage to stream, on average, more than 3 sessions a day.

Comparing that professional gaming production to that of YouTube emphasizes, again, how different those contents are: professional YouTubers such as Mistery Guitar Man [5] or Smosh [6] only release a few videos per week, whereas 30% of the popular Starcraft II players in our dataset broadcast their games in at least two sessions a day. In that regard, gaming can be said easier than producing YouTube content.

E. Understating the spectator

In this section we try to characterize the spectator behavior as we did with the streamer's.

[5]https://www.youtube.com/user/MysteryGuitarMan
[6]https://www.youtube.com/user/smosh

Sojourns	$20,105,627$	
Type	Median	%
$ON \approx ON$	7.5	82.05
$ON \approx OD$	48.5	5.66
$ON \approx OFF$	189.8	2.94
$OD \approx ON$	115.22	0.04
$OD \approx OD$	1.4	0.65
$OD \approx OFF$	40.20	0.05
$OFF \approx ON$	102.91	0.52
$OFF \approx OD$	175.6	0.03
$OFF \approx OFF$	3.18	8.06

Table III: General sojourn percentages by kind.

1) Channel-surfing: We define channel-surfing as a quick sojourn of a spectator in a streamer channel. Such behavior was clearly observed in our dataset. We evaluated more than 20 million sojourns and Table III shows the percentage of each type (defined as a pair entry and exit states). The first column shows the states of the streamer when the spectator joined and left his or her channel. The second column shows the median time of the sojourn. The last column displays the percentage of each kind of sojourn.

The sum of $ON \approx ON$, $OFF \approx OFF$ and $OD \approx OD$ sojourn percentages are equal to 90.8% of the total. Their duration medians are, respectively, 7.5, 3.18 and 1.4 minutes. Despite that, about 20% of the $ON \approx ON$ sojourns and 30% of the $OFF \approx OFF$ sojourns lasted less than a minute. These data indicate that this is a significant behavioral pattern among spectators, and suggest that spectators use Twitch.tv without knowing *a priori* the content they want to consume.

The $OD \approx OD$ sojourn is associated with the shortest duration median (1.4 minutes), which can be explained by the own meaning of the state. When someone joins a channel and the streamer is finishing the transmission, spectators will tend to leave, as they won't be able to see the streamer playing. The $OFF \approx OFF$ sojourn small median also has an obvious interpretation. After joining the streamer channel and finding out that he or she is offline, the spectator quickly leaves the stream.

2) Channel switching: Each spectator seeks to consume a content that pleases him or herself. While watching a given channel, a spectator may decide to switch to another channel in order to find a content that better suits his or her tastes. In this scenario of channel switching, it is important to understand the context where switching happens and what it does mean.

To achieve that, we analyzed our dataset and found exactly 2,386,972 switches during the 139 days during which the data was collected. We summarized the types of switches in Table IV. The first column shows the states of the streamers at origin and destination. The second column shows the median of the spectators sojourn time before making the switch - we used the median because the average sojourn

Kind	Sojourn duration	#	%
$ON \rightarrow ON$	7	1,722,321	72%
$OD \rightarrow ON$	27	292,045	12%
$OFF \rightarrow ON$	20	180,874	8%
$ON \rightarrow OFF$	10	59,781	3%
$OFF \rightarrow OFF$	9	54,864	2%
$ON \rightarrow OD$	6	47,402	2%
$OD \rightarrow OFF$	30	10,156	0%
$OD \rightarrow OD$	15	9,473	0%
$OFF \rightarrow OD$	12	4,817	0%
Total:	—	2,386,972	100%

Table IV: General data about channel switching.

time is distorted by extreme values. The absolute number of channel switches (#) and the corresponding percentage (%) are displayed in the last two columns.

We believe that the state the streamer is when the channel switching happens is related to the reason behind the switch. Understanding this reason is important to characterize how and when streamers attract their spectators, and in order to determine how popular a streamer is - which is extremely interesting in the streamer's perspective, since they profit from being popular [5].

For example, a possible interpretation of the $ON \rightarrow X$ switch is that an spectator leaves a session that is not satisfying and goes to another streamer channel. A good analogy to such an action is television: if a spectator is watching a show and in the middle of it changes to another channel, he is probably not enjoying the show very much. Notice that the largest number of channel switching is from the $ON \rightarrow X$ kind, what is explained by the visibility that popular streamers channels achieve in Twitch.tv recommendation system.

To support our hypothesis, in Fig. 6, we plotted a graph comparing, in an $ON \rightarrow ON$ channel switch, how long does a spectator stay in each of the channels. Notice that the spectators tend to stay slightly longer watching the second streamer, supporting the argument that when an $ON \rightarrow X$ happens, the spectator is not satisfied with the content that the first streamer is streaming.

The $OD \rightarrow X$ is the second more popular kind of channel switch - confirming the high spectator evasion described in Section V-C. In this case, assuming that there is a preference for the second streamer does not make any sense from a semantic perspective, once the spectator might be leaving the channel because he or she will not be able to consume its content anymore. This same statement is also valid for switches from $OFF \rightarrow X$ kind, where a spectator, unable to consume the content of one channel, switch to another one.

3) Spectator assiduity: We here define and measure how assiduous the spectators are to the streamers they watch. By "assiduity", we mean a propensity for a spectator e to stick to a small number of streamers among the set S of all the

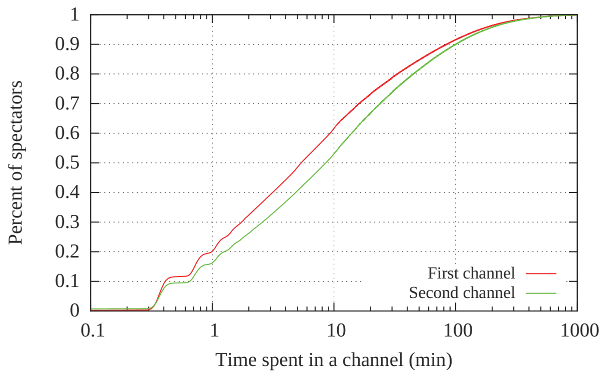

Figure 6: Spectators spend more time in a channel after a switching.

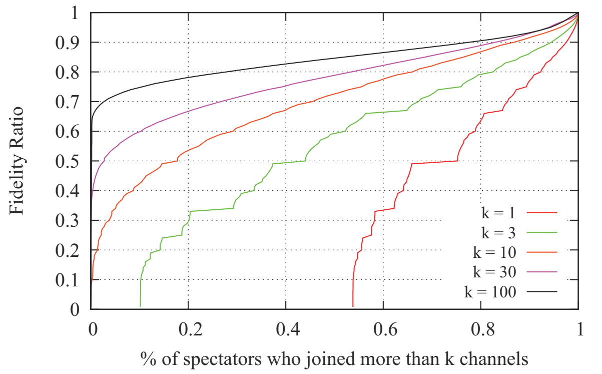

Figure 7: Cumulative distributions of the fidelity ratios of all spectators who joined at least k times.

streams studied in this article. More formally, and noting $incs(e, s)$ the number of times the spectator e joined the channel of a streamer s, the following function F, called *fidelity ratio*, is here proposed:

$$F : e \mapsto 1 - \frac{|\{s \in S \mid incs(e, s) \neq 0\}|}{\sum_{s \in S} incs(e, s)} \tag{1}$$

F maps every spectator e to a number in $[0; 1)$. The higher $F(e)$, the more assiduous the spectator e. The metric is obviously not trustworthy for spectators with few visits to Twitch.tv. In particular, a spectator with one single connection necessarily has a null fidelity ratio.

In Fig. 7, each curve relates to a value k and shows the cumulative distribution of the fidelity ratios of all spectators who joined at least k channels. Those curves confirm what the intuition tells: the more channels joined, the more assiduous the spectator, i.e., the spectators coming again and again on Twitch.com do so to always watch the same streamers. What the figure does not show is that most spectators join very few channels. Among the $1,490,121$ spectators, $455,009$ ($\approx 30\%$) joined more than 3 channels

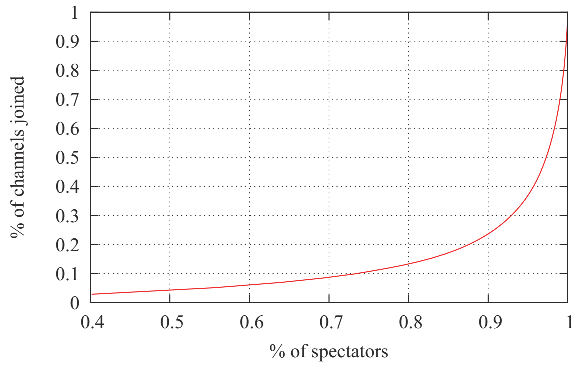

Figure 8: Cumulative distribution of the connections to all channels over the spectators.

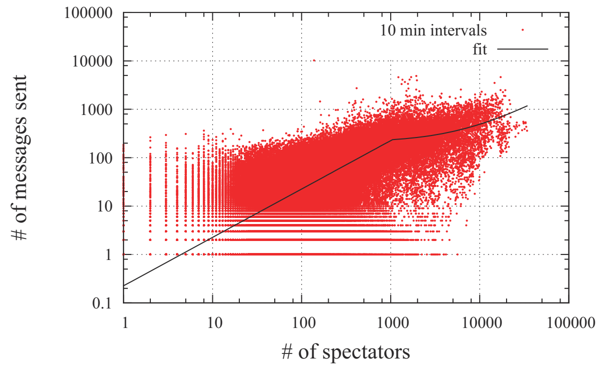

Figure 9: Cumulative distribution of spectators per fidelity ratio.

and only $279,005$ ($\approx 1.87\%$) joined more than 10 channels.

Fig. 8 shows the cumulative distribution of the connections to all channels over the spectators. The top 10% of the spectators are responsible for almost 80% of the connections. More generally, most of the content streamed from Twitch.tv is watched by a small group of passionate spectators. The same small group that was shown earlier to be assiduous (high k value).

4) Chatting: We analyze here the relationship between the number of spectators logged into a Twitch.tv chat and the activity of that chat, i.e., how fast the conversation. More precisely, the activity y is here measured as the number of messages sent in a 10 minutes interval and the related number of spectators x is taken at the end of that time period. We consider all ten minutes intervals in all the channels we collected.

The relationship between the number of spectators (explanatory variable) and the activity (response variable) can be modeled by regression. The curve we propose to fit is composed of two straight lines. The first one starts at $(0;0)$ (because a chat with no spectator cannot have activity) and, at some point, the second line starts where the first one ended. Mathematically, here is the model (where three parameters need to be estimated: a, b and c):

$$y = \begin{cases} a \times x & \text{if } x < c \\ b \times x + (a - b) \times c & \text{if } x \geq c \end{cases} \quad (2)$$

The intuition behind that model is that after certain number c of spectators logged into the chat and a related high chatting activity ac, it becomes harder to communicate and the activity enters a second regime where the growth of the activity w.r.t. the number of spectators is not as intense ($b < a$).

The breaking point c is estimated by an exhaustive search in $[1; 1000]$. For each such estimation, a minimization of the squared residuals allows to fit the first line to the points with abscissas smaller than c and the same method is used to fit the second line to the points with abscissas greater or equal to c. The overall sum of the squared residuals conditions the choice the best estimations for c: the smaller the sum, the better. Fig. 9 shows all data points and the model that fits them. The change of regime happens at approximately $c = 1,000$ spectators. At that abscissa, there are $ac = 22.5$ messages that are sent every minute.

To validate the model, its *rms of residuals* is compared to a more naive model that passes by $(0;0)$ and has the same number of parameters: $y = ax^3 + bx^2 + cx$. The model we proposed is better: its *rms of residuals* is 50.63, whereas that of the cubic fit is 53.06.

VI. Conclusion

In this paper we presented a generic model that can be used to analyze video game live-streaming and used it in order to characterize the Starcraft II sub-community (which is mainly a part of the e-sport sub-community). We analyzed data from Twitch.tv, gathered using the website's own API and an IRC crawler. This paper has shown, among other results that: (i) spectators have clear behavioral patterns such as channel surfing and leaving quickly close to the end of a streaming session, (ii) the content produced is longer and less edited than most of online entertainment content, (iii) there is semantic meaning to channel switching (iv) the content is mainly consumed by a small fraction of very assiduous streamers, (v) we can predict the number of messages sent in the chat using a closed formula. Those results are of major interest for all members of the gaming community and also for the scientific one.

As future work, we intend to extend our model to other games that are being broadcasted in Twitch.tv and also check its applicability to other media, such as Youtube. We also intend to design, implement and evaluate novel algorithms that exploit the knowledge of our characterization provides about the behavior of both streamers and spectators, such as recommending streamers in real time.

Acknowledgment

This work was partially supported by CNPq, CAPES, FAPEMIG, and InWEB.

References

[1] M. Kaytoue, A. Silva, L. Cerf, W. Meira, Jr., and C. Raïssi, "Watch me playing, i am a professional: A first study on video game live streaming," in *Proceedings of the 21st International Conference Companion on World Wide Web*, ser. WWW '12 Companion. New York, NY, USA: ACM, 2012, pp. 1181–1188. [Online]. Available: http://doi.acm.org/10.1145/2187980.2188259

[2] (2014-05-15) Twitch Named Best Games Related Website for 3rd Consecutive Year in 18th Annual Webby Awards. [Online]. Available: http://www.businesswire.com/news/home/20140429006770/en#.U3UikPgc3Vs

[3] (2014-05-15) How Big Is Twitch's Audience? Huge. [Online]. Available: http://www.forbes.com/sites/davidewalt/2014/01/16/twitch-streaming-video-audience-growth/

[4] T. L. Taylor, *Raising the Stakes : E-Sports and the Professionalization of Computer Gaming*. MIT Press, 2012.

[5] T. Smith, M. Obrist, and P. Wright, "Live-streaming changes the (video) game," in *Proceedings of the 11th European Conference on Interactive TV and Video*, ser. EuroITV '13. New York, NY, USA: ACM, 2013, pp. 131–138. [Online]. Available: http://doi.acm.org/10.1145/2465958.2465971

[6] (2014-2-05) Twitch is 4th in Peak US Internet Traffic. [Online]. Available: http://blog.twitch.tv/2014/02/twitch-community-4th-in-peak-us-internet-traffic/

[7] G. Cheung and J. Huang, "Starcraft from the stands: Understanding the game spectator," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 763–772. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979053

[8] K. Sripanidkulchai, B. Maggs, and H. Zhang, "An analysis of live streaming workloads on the internet," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 41–54. [Online]. Available: http://doi.acm.org/10.1145/1028788.1028795

[9] J. D. Weisz, S. Kiesler, H. Zhang, Y. Ren, R. E. Kraut, and J. A. Konstan, "Watching together: integrating text chat with video," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '07. New York, NY, USA: ACM, 2007, pp. 877–886. [Online]. Available: http://doi.acm.org/10.1145/1240624.1240756

[10] D. A. Shamma, L. Kennedy, and E. F. Churchill, "Watching and talking: media content as social nexus," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, ser. ICMR '12. New York, NY, USA: ACM, 2012, pp. 12:1–12:8. [Online]. Available: http://doi.acm.org/10.1145/2324796.2324811

[11] P. Cesar and D. Geerts, "Past, present, and future of social tv: A categorization," in *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, Jan 2011, pp. 347–351.

[12] L. Oehlberg, N. Ducheneaut, J. D. Thornton, R. J. Moore, and E. Nickell, "Social tv: Designing for distributed, sociable television viewing," in *Proc. EuroITV*, vol. 2006, 2006, pp. 25–26.

[13] S. R. Foster, S. Esper, and W. G. Griswold, "From competition to metacognition: designing diverse, sustainable educational games," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 99–108.

[14] J. J. Thompson, M. R. Blair, L. Chen, and A. J. Henrey, "Video game telemetry as a critical tool in the study of complex skill learning," *PLoS ONE*, no. 9, 2013. [Online]. Available: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0075129

[15] (2014-03-10) YouTube statistics. [Online]. Available: http://www.sysomos.com/reports/youtube/