Live streaming: A New Haven for Advertisement Fraud

Andrew Swindlehurst – Data Analyst at PPC Protect

Abstract - Live streaming has exploded in popularity over the last decade, and with it, a new advertising medium has arisen. Opportunistic individuals have used the growth to defraud platforms and advertisers, using viewbotting services to artificially inflate audience numbers and fraudulently consume advertiser's budgets to increase their revenue. Some platforms are taking actions against viewbotting services, however, to protect advertisers, more proactive measures are required. Through literature review, a number of topics are covered, such as the motivations of the users of viewbotting services, the ineffectiveness of reactive litigation, and some suggested proactive countermeasures that could be implemented to live streaming platforms.

Introduction

Live streaming is the act of simultaneously recording and broadcasting media to an audience in real time, and although not a new idea, live streaming has exploded in popularity over the last decade. Live streaming's growth has coincided with the creation of some notable live streaming platforms such as Justin.tv, now Twitch.tv, YouTube's live streaming service, and Ustream. Two of which feature on Alexa's top 50 global sites with YouTube at position two, behind Google, and Twitch.tv at position 32, ahead of LinkedIn, eBay, and Bing (Webarchive.org, 2018).

As these platforms grew into the multibillion-dollar industry it is today; it became clear that it was a prime advertising space. Platforms such as Twitch.tv and YouTube's Gaming service have a highly focused demographic. Twitch.tv released figures showing their audience is 81.5% Male, 55% aged between 18-34 (Twitch Advertising, 2018) and all incredibly passionate about video games. This millennial demographic has traditionally been very difficult to reach and has previously puzzled marketers (Forbes, 2016). So, with over 15 million daily viewers, if you're looking to advertise to male millennials, live streaming would appear to be an excellent place to start.

However, as with anything, some people are willing to bend or break the rules to get ahead of the competition. For live streaming, this comes in the form of "viewbotting". Viewbotting is a method of artificially inflating viewer count, and is against most, if not all, live streaming platform's terms and conditions. Similar to clickbots and click farms in "classical" digital marketing, these tools are used to defraud advertisers, viewers, or both.

Through a literature review, we aim to explore the motivations behind viewbotting (malicious or otherwise), the problems caused by viewbotting, what is being done to prevent viewbotting, and other possible methods of preventing viewbotting. Given that Twitch.tv is the largest platform purely dedicated to live streaming, it will be the primary target for this research.

Motivations for Viewbotting

Twitch.tv has a "most popular first" approach to displaying the current active live streams; resulting in the games and streamers with the most current viewers always being the first to be seen. This approach results in a barrier to entry when beginning to grow a genuine audience. Viewbotting services provide a simple and effective method of overcoming this barrier to entry; one service is even going

as far as using the tagline “Get Popular Now” (Twitch Buddy, 2018).

Growing an Audience

Growing your live stream audience is a commonly occurring theme among viewbotting services. The issue faced by a live stream with lower viewer count is that they will rarely be seen by anyone casually browsing the website, as they are hidden under larger streams. To get the exposure you need an audience, and a cursory search on an internet search engine will quickly point you towards a viewbotting software to start building an audience. The software fraudulently inflates the live stream viewer count by using a botnet to simulate live stream viewers.

With the increase in viewer count, Twitch.tv will list the particular stream higher, and thus the live stream will enter the previously mentioned cycle of exposure. Once the live stream has picked up a large genuine audience, the viewbotting software can be stopped, as at this point, the live stream has enough traction to be self-sustained.

Malicious Intent Against Other Live Streams

Viewbotting services are non-discriminatory on what channel the viewbotting service is targeting. This can pose an issue for legitimate live streamers as they can be anonymously targeted with a viewbotting service putting their live stream at risk of being banned for breaking the terms of service.

Using viewbots in an attempt to coax action out of the live streaming platforms is so prevalent Twitch.tv has a dedicated page helping people targeted by viewbotting services (Twitch, 2016). A notable example is the League of Legends player Søren “Bjergsen” Bjerg, who took to Twitter after being notified of potentially being viewbotted (Twitter.com, 2018).

Affiliates, Sponsorships & Partnerships

Some live streaming platforms allow for monthly payments to support a live stream, sometimes known as “subscriptions”. Twitch.tv, for example, has multiple subscription options, starting at only 5 USD per month. To be eligible for monetisation via a subscription model on Twitch.tv, the live stream must be a “Twitch Partner”. One of the requirements for applying for partnership is to have “an established and steadily growing audience and chat” (Twitch, 2018), and some viewbotting services target this as an advertising standpoint (Twitch Buddy, 2018).

Becoming a “Partner” is not the only way that viewbotting services can be used to defraud companies. Live streamers are regularly sponsored by companies to advertise their products or brands, sometimes in return for monetary rewards. A common occurrence is “Sponsored Streams” where live streams are paid to play a specific game for a set amount of time. If, for example, a game studio approached a streamer with an average of 3,000 viewers and paid the streamer a sum of money to stream their game for a set amount of time. The game studio would expect their game to be exposed to 3,000 genuine viewers, however, if the live streamer is using a viewbotting service the game studio would only be exposing their game to a fraction of the number of people they initially believed.

Desperation

A notable event involving viewbotting was *The Attack*’s Kevin Pereira, who admitted the show was using a viewbotting service to keep the show afloat. In an interview with Polygon, Pereira claimed that “with people’s jobs on the line, he was feeling desperate to try and keep the show going to keep as many people employed as possible.”

Pereira closed *The Attack*’s Twitch channel before it was possible for Twitch.tv (the

live streaming platform used to host the show) to take action against it and Pereira has acknowledged “he knew he was doing something wrong”. Pereira also stated “So instead of trying to make the content better or refocus my strategy, with the limited time we had left, I decided to shortcut it and try to get some extra views on the channel.”

Although Pereira did appear to have good intentions stating “I tried to save jobs, I really did, and hope that in the end that I didn’t do any irreparable damage to people’s careers.” it does not change how problematic his actions are for advertisers.

Twitch.tv’s Litigation Attempts

Some attempts are being made to curb the use of viewbotting services, notably Twitch.tv suing seven different viewbotting services (Polygon, 2018). A result from one of these cases netted Twitch.tv $55,000 in statutory damages, and a further $1,316,139, representing the profits earned through the sale of their services along with an order for the viewbotting service to cease operation (Chalk, 2018).

Litigation, though a start, is a slow process with many issues. Litigation may be levelled against larger sites, as Twitch has done, however, it is near impossible to eradicate all viewbotting services. The difficulty of the task is due to the ease in which new viewbotting services can startup and begin selling their products, some anonymously, going as far as using cryptocurrency making it a costly undertaking to identify individuals to litigate against.

While live streaming platforms are playing *Whack-a-Mole* with viewbotting services, ad spend is being burned on artificial audiences. A more comprehensive and proactive approach is required for removing viewbots before a large amount of damage is caused to advertiser budgets.

Proactive Prevention

An example of a live stream with a fraudulent audience can be seen in figure 1. This live stream has been banned by Twitch.tv for viewbotting in 2015 and when compared to a genuine audience of a similar size, the difference is stark. The viewbotted audience joined the stream at a much more rapid pace, gaining approximately 400 viewers in only 15 minutes. The rapid growth of the fraudulent audience is over seven times more than the genuine live stream gained within a 15 minute period. This style of viewbotting has since been *mostly* phased out as it is relatively simple to spot such a significant jump in viewer count. Viewbotting services have since become more discreet in their actions, making them harder to detect.

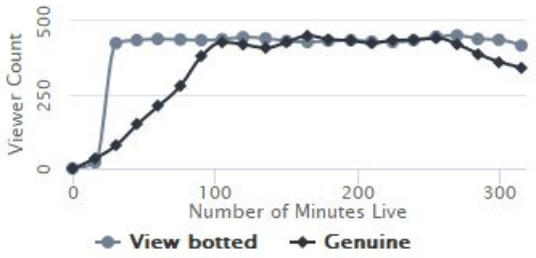

Figure 1. Comparison between two live streams of similar size. One with a genuine audience, and the other a viewbotted audience.

The litigation measures taken by Twitch.tv may be the first steps in preventing viewbotting operations from operating long term, but it does little to prevent private operations or short-term business plans. To prevent such business ventures, a more robust solution is required. Proposals by C.A.Watts and N. Shah may be such solutions.

The Watts Solution

Watt's proposal uses Deep Ensemble Recurrent Artificial Neural Networks, as well as a variety of other techniques (Watts, 2016). He attempts to identify preventable viewbotting, through the use of a number of common and successful deep learning tools, Watts aimed to break down live stream traffic into two categories, real and fake viewers, with a high degree of accuracy. Some of the analysis tools he used were Naïve Bayes, Support Vector Machines, K-Means, Random Forests and Deep Ensemble Recurrent Artificial Neural Networks (DERANN).

Watts has also listed out a number of components he intended to use to identify fake viewers, some these identifiers include:

- Chat to viewer ratio
- Viewer to follower ratio
- Moving average of the number of viewers
- Derivative of the number of viewers.
- Names of chatters
- Moving average of chat volume per chatter

Chat to viewer and viewer to follower ratios are simple to identify. To access the chat on Twitch.tv and the follow function a user must register an account. Most viewbotting services do not register accounts preventing them from chatting and following. The registered viewer to unregistered viewer ratio is expected to be relatively constant across the whole website. A particular live stream with a large number of unregistered viewers could be a sign of viewbot activity on that particular live stream.

If a viewbot is sophisticated enough to use registered accounts, a username will be required. Viewbotting platforms would need to programmatically generate usernames to chat, Watt's proposed using a measure of clustered similarity to identify programmatically generated account names. In conjunction with the names of users chatting, viewbotters can also be identified if there is a substantial drop in the average message rate per viewer in a live streams chat. This would indicate a large number of viewers connecting to the chat and not chatting.

The rate of change of viewers on a live stream, or the derivative of number of viewers, is useful to spot viewbotting as covered with Fig. 1. It should be mentioned that large spikes in viewers are not necessarily a sign of viewbotting, as live streamers sometimes "host" other live streams on their own and some direct their viewers to other streams resulting in a large boost of viewers in a small amount of time. Differentiating referral viewer spikes is possible when combined with the other identifiers. Identifiers such as chat to viewer ratio or the moving average of chat volume per chat user, as most viewers following the referral tend to be active in chat.

With the identification parameters laid out, the algorithm was trained using a live stream with a large, established audience and on a control live stream with a varying viewership percentage of viewbots. After training the algorithm was scaled up to be able to process a large number of live streams in parallel.

According to Watts, he "believes that the system will be able to effectively monitor all streams of sufficient science." and "the system is able to monitor 64% of the 1.09 million Twitch.tv users".

The Shah Solution

Unlike Watts, Shah proposes an offline solution focusing on aggregated behaviours (Shah, 2018). Shah's solution aims to build a model of *normal* viewing behaviours by observing behaviours in aggregate. He would then identify behaviours that stand out from the model of *normal* aggregate behaviour. Using this method of building a *normal* aggregate

viewing model, identifying viewbotting would be similar to an outlier detection problem.

Modelling Broadcasting Behaviour proved to be difficult given the lack of streams by temporal features of their constituent views. For each viewer on a live stream, Shah's solution is interested in an individual viewers start and end time, as these are near impossible to spoof while adhering to the goals of a viewbot.

Factoring a viewer's start and end time against the overall length of a live stream along with distribution brackets, Shah constructed a *normal* viewing behaviour model. To differentiate between authentic and viewbotted broadcasts requires an outlier-detection test in which the interest is abnormal broadcasts. Shah accomplished this by identifying a way to measure deviance between the broadcast distribution, its associated bracket distribution, and identifying a classification threshold to set for the resulting deviance scores.

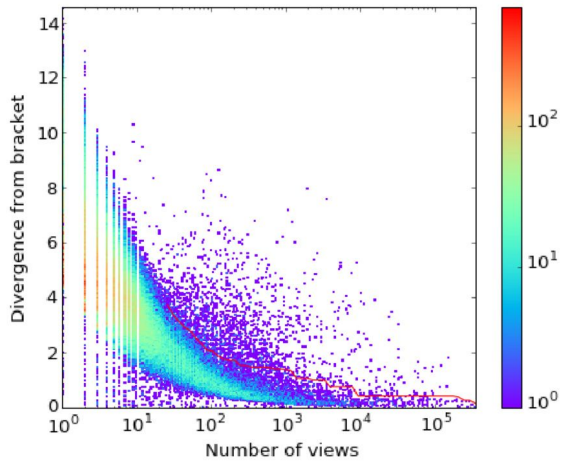

Figure 2. Each broadcast is a point with view count on the x-axis, deviance between broadcast/bracket on the y-axis and area density denoted by colour. The red line indicates the decision boundary.

Figure 2 shows high variance in deviance value for *lower* view live streams, and that variance drops as the audience grows larger. The broadcasts above the red line in figure 2 are broadcasts that Shah suspects ground-truth labelled data (or provable data), so an unsupervised model is used. Instead of focusing on engagement based identifiers like Watts, Shah instead focuses on modelling live

contain viewbots and thus uses those broadcasts to identify viewbots. An important step Shah took is to distinguish between authentic and viewbotted views.

Shah's results indicated 98% of outlier broadcasts to be viewbotted, 99% of non-outlier broadcasts not to be viewbotted, and over 90% precision in identifying views in large viewbot attacks. This degree of accuracy legitimises the proposed unsupervised classification approach.

Conclusion

Live streaming is a booming industry, coming to the forefront of online entertainment in the last half-decade. Occupying two places in the top 50 websites, Live streaming's growth has been explosive; although live streaming has experienced growing pains. Viewbotting has become a plague on the live streaming that live streaming platforms have released guides on how to deal with being maliciously viewbotted.

Some platforms have started taking action such as Twitch.tv's lawsuit against multiple popular viewbotting services. Using lawsuits to shut down viewbotting platforms does not protect advertisers. A more proactive solution with rapid removal of exploitative users is required. Watts and Shah both provide such solutions with Watt's Deep Ensemble Recurrent Artificial Neural Networks, and Shah's offline solution focusing on aggregated behaviours. Both of these solutions could highlight fraudulent viewership daily, rather than a large amount of time it takes for litigation to take effect.

References

Web.archive.org. (2018). [online] Available at: https://web.archive.org/web/20150302173920/h [Accessed 29 Jun. 2018].

Twitch Advertising. (2018). *Audience*. [online] Available at: http://twitchadvertising.tv/audience/ [Accessed 29 Jun. 2018].

Twitch-buddy, (2018). [online] Available at: http://www.twitch-buddy.com/index [Accessed 9 Jul. 2018].

Twitch. (2016). *How to Handle View/Follow-Bots*. [online] Available at: https://help.twitch.tv/customer/portal/articles/2435640 [Accessed 9 Jul. 2018].

Twitter.com. (2018). *Twitter*. [online] Available at: https://twitter.com/Bjergsen/status/724530907030265856 [Accessed 9 Jul. 2018].

Twitch. (2018). *Twitch*. [online] Available at: https://www.twitch.tv/partner/signup [Accessed 9 Jul. 2018].

Polygon. (2018). *Twitch crackdown on 'view-bots' includes a lawsuit against seven of their makers*. [online] Available at: https://www.polygon.com/2016/6/18/11968762/twitch-viewbots-lawsuit [Accessed 9 Jul. 2018].

Chalk, A. (2018). *Twitch view bot makers ordered to pay nearly $1.4 million in lawsuit loss*. [online] pcgamer. Available at: https://www.pcgamer.com/twitch-viewbot-makers-ordered-to-pay-nearly-14-million-in-lawsuit-loss/ [Accessed 9 Jul. 2018].

Watts, C. (2016). *IDENTIFYING POPULARITY MANIPULATION OF LIVESTREAMS ON TWITCH.TV*. B.S. The University of Georgia.

Shah, N. (2018). FLOCK: Combating Astroturfing on Livestreaming Platform. In: *WWW '17 Proceedings of the 26th International Conference on World Wide Web*. [online] Pittsburgh, pp.1083-1091. Available at: http://www.www2017.com.au/ [Accessed 9 Jul. 2018].