

Investigating VPN Ads on YouTube

Omer Akgul, Richard Roberts, Moses Namara*, Dave Levin, Michelle L. Mazurek
University of Maryland
Clemson University*

Abstract—VPN companies commonly advertise VPNs on YouTube through influencer marketing techniques. In this proposal we describe how we obtained a dataset of YouTube videos that contain VPN ads and outline our analysis plan.

I. INTRODUCTION

VPNs are an increasingly popular end-user privacy solution [4], [5]. VPN companies commonly advertise their products with influencer marketing techniques on YouTube¹.

With influencer marketing, the content producer (called *YouTubers* on YouTube) promotes products in exchange for direct compensation and/or a commission from associated sales. We notice that YouTubers seem make a wide variety of claims when promoting VPNs on YouTube, which include vague and in some cases potentially misleading statements about the capability of VPNs and internet threats in general².

In this proposal, we first describe how we measure the prevalence of videos on YouTube that contain *influencer VPN ads* (IVAs). We then outline and seek feedback on our analysis plan to better understand user impact.

II. THE DATASET

We use random prefix sampling [6], which theoretically allows us to sample 1/64th (1.6%) of all videos on YouTube. In practice, our implementation attempted to sample 1.5% of all videos, capturing the brief details (view count, date, etc.) of 86,279,708 videos. We collect additional details and (when available) English subtitles for 10,697,546 of the most viewed videos (views ≥ 800), covering $> 99\%$ of total views in the initial dataset.

By comparing our data to a release from YouTube [1], we find that we likely obtained 95.8% of all possible videos that could have been obtained with our implementation. Overall, we likely sampled of 1.4% of all YouTube videos.

To identify videos containing IVAs, we first identify all videos whose English subtitles include the word “VPN” (1,751 in total) to establish a *candidate dataset*. Next, three researchers applied an open-coding approach on $\sim 25\%$ (425 videos) of the dataset to precisely define what IVAs are. The final definition includes all videos that mention VPNs and have affiliations with VPN companies. Three researchers independently applied this definition to adjudicate an additional 175 videos ($\sim 10\%$ of the candidate dataset), reaching

Krippendorff’s α of 0.852; therefore, validating this definition [3]. Finally, all three researchers split and adjudicated the remaining candidate videos. In total, we found 359 videos that contain IVAs in our dataset ($4.1 * 10^{-4}\%$ of all videos). We call this final sample *the random sample*.

This suggests that $\sim 25,200$ videos (Agresti-Coull [2] $CI_{95\%} = (22,696 - 28,016)$ ³) contain IVAs on YouTube.

To further supplement our dataset, we gather the details and subtitles of all videos related to⁴ videos in the random sample (3,043 candidate videos) or belonging to one of the 241 channels that produced the random sample (14,969 candidate videos). Data was collected between August-December, 2020.

III. ANALYSIS

The analysis plan consists of several stages:

a) Categorizing ad content: We will qualitatively code the content of the videos in the random sample to identify several properties:

- Threat model: The relationship between adversaries, attacks, and consumer assets YouTubers describe.
- Product features: Commonly advertised features include: many servers, having “military-grade encryption”, etc.
- Conditions VPN’s should be used under: public WiFi’s, when traveling, torrenting, searching QAnon posts, etc.

b) Analysis of channels/YouTubers: We will analyze and describe what types of channels promote VPNs and how their promotion approach relates to their audiences. A qualitative coding approach could be appropriate, in addition to data obtained through the YouTube API.

c) Content analysis over time: How has the content of IVAs changed over time? How did their prevalence change? We plan to use videos within the same channels to reduce the noise of time series data.

d) How do VPN companies factor in?: Sponsoring VPN companies provide YouTubers guidelines on how to advertise. We plan to compare these guidelines to the final products.

e) User study: We will design and execute a user study with the goal to identify a set of videos that embody specific features and showing them to participants, measuring how their mental models change after watching.

Key challenges include identifying good metrics for mental models and matching participants to the intended audiences of various YouTube channels.

¹e.g., <https://nordvpn.com/influencers/>, <https://expressvpn.com/influencers>

²Two examples of unusual ads/claims include https://youtu.be/PWg_-Fjja3Y?t=40, and <https://youtu.be/CwF8DYf5dDc?t=673>. Please note that the second link contains adult language.

³The CI does not account for human errors in labeling.

⁴as determined by the YouTube data api: <https://developers.google.com/youtube/v3/docs/search/list#relatedToVideoId>

IV. ACKNOWLEDGEMENTS

This research was supported in part by a UMIACS contract under the partnership between the University of Maryland and DoD.

REFERENCES

- [1] Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*, 2016.
- [2] Lawrence D Brown, T Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical science*, pages 101–117, 2001.
- [3] Klaus Krippendorff. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research*, 30(3):411–433, 2004.
- [4] Moses Namara, Daricia Wilkinson, Kelly Caine, and Bart P Knijnenburg. Emotional and practical considerations towards the adoption and abandonment of vpns as a privacy-enhancing technology. *Proceedings on Privacy Enhancing Technologies*, 2020(1):83–102, 2020.
- [5] Goddy Ray. Guide to vpn users statistics for 2020, Nov 2020. Available at <https://surfshark.com/blog/vpn-users>.
- [6] Jia Zhou, Yanhua Li, Vijay Kumar Adhikari, and Zhi-Li Zhang. Counting youtube videos via random prefix sampling. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 371–380, 2011.