Investigating the VPN Recommendation Ecosystem

Reethika Ramesh
University of Michigan

Armin Huremagic
University of Michigan

Chad Sharp
University of Michigan

Roya Ensafi
University of Michigan

Abstract—**Commercial VPNs are multi-billion dollar industry with hundreds of VPN providers from which users can choose. Consequently, there exists a plethora of websites that consistently churn out recommendations and rankings of VPN providers. Alarmingly, from our prior work with VPN providers, we learned that these websites are largely money-motivated: they often auction-off top spots, contain bogus rankings, and are affiliated or even owned by VPN companies. Our project will conduct data-driven investigation into the VPN recommendation ecosystem to uncover patterns of biased recommendations.**

I. INTRODUCTION

The use of Virtual Private Networks (VPNs) has been growing rapidly due to increased public awareness of online risks, expanding geographic restrictions, and aggressive marketing campaigns of VPN products [1]–[5]. The commercial VPN industry has merited a handful of academic investigations, but has 100s of providers from which users can choose [6]–[9]. As with any market having abundant choices, there is a plethora of resources and websites that constantly churn out rankings of VPN providers based on various, often inconsistent heuristics.

In our previous work, we interviewed nine VPN providers to understand the issues and challenges in the VPN ecosystem from their point of view. Alarmingly we learned that these recommendation websites are often money-motivated, contain paid review spots and rankings, and are sometimes even owned by VPN companies. Additionally, a handful of investigative reports revealed that VPNs and their parent companies own many recommendation websites [10]–[13].

Since these recommendation sites can influence which VPNs users buy, and often serve as a user's first point of research, shedding light on this ecosystem is critical to ensure users are not exploited. In this proposal, we describe how we will investigate this ecosystem by conducting measurements and creating an extensive dataset of VPN recommendation websites. We will describe the various aspects of these websites that we will compare and analyze for similarities and uncover patterns of biased recommendations.

II. DATA COLLECTION

We begin by independently creating a list of intuitive search terms that a user would use to search for VPNs. Then, we iteratively collect *related* search terms using popular search engines such as Google, Bing, DuckDuckGo. Through this process, we collect 34 key words that users use on search engines while exploring for a VPN to use.

Using this list of key words we programmatically query three popular search engines (Google, Bing, DuckDuckGo), to collect a corpus of VPN recommendation sites. For each keyword, we record the top 100 search results. We sanitize this list of web pages to exclude VPN providers' own websites, links to VPNs on app stores (Google Play, Chrome App Store, Mac App Store) and results from user posts on popular forums like Reddit, Quora, Wikipedia, and YouTube.

Finally, for each website on the final list, we conduct measurements to collect all the resources, scripts, and links loaded by each along with a complete copy of the DOM.

III. DATA CHARACTERIZATION & ANALYSIS

Using our measurements, we create an extensive dataset of VPN recommendation websites, and study the ecosystem by analyzing different aspects of these sites to reason about the objectivity of their recommendations. So far, we have collected 2,934 unique web pages by querying 34 key words on all three search engines. Our analysis plan spans the following stages:

1) *Extract the meaningful data*: We will use Python's Beautiful Soup library [14] to extract useful text and remove boilerplate data. We will then run a script to extract the VPNs and accompanying text from the webpage.

2) *Compare VPN rankings between recommendation sites*: We hypothesize that a set of VPN providers invest heavily in affiliate marketing and buy review spots across various recommendation websites. We will investigate this using measures for comparing ranked lists such as Rank-Biased Overlap (RBO) [15] and Kendall's τ [16]: if a set of n VPN providers invest in such methods, the top n spots on different recommendation websites for multiple key words will be comparatively more stable and consistent than the rest of the rankings. This stability may also decrease with popularity (search result rankings) of the recommendation sites. We will compare rankings with trusted recommenders [17], [18], and also combine insights with the following analyses to uncover commonalities and potential biases in recommendations.

3) *Content analysis*: We will compare *homogeneity in content* for each VPN provider (i.e. comparison heuristics, pros and cons *etc*) across different websites to uncover possible common relationships between the websites. To assess *webpage similarity*, we will compare the DOM structures, scripts and other resources including images that are loaded on each site.

4) *Analyze links on the websites*: We notice that recommendation sites usually contain links to the VPN provider for each ranking. We will collect and analyze these links, capture each redirect, and characterize if and how affiliate links and intermediaries play a role.

REFERENCES

[1] A. McDonald, M. Bernhard, L. Valenta, B. VanderSloot, W. Scott, N. Sullivan, J. A. Halderman, and R. Ensafi, "403 Forbidden: A global view of CDN geoblocking," in *Proceedings of the 2018 Internet Measurement Conference*, 2018.

[2] C. Cimpanu, "Data breach at Russian ISP impacts 8.7 million customers," ZDNet, 2019, https://www.zdnet.com/article/data-breach-at-russian-isp-impacts-8-7-million-customers/.

[3] M. Snider, "ISPs can now collect and sell your data: What to know about Internet privacy rules," USA TODAY, 2017, https://www.usatoday.com/story/tech/news/2017/04/04/isps-can-now-collect-and-sell-your-data-what-know-internet-privacy/100015356/.

[4] I. Bulatovaitė, "Your master guide to VPN usage statistics," https://surfshark.com/blog/vpn-users, 2020.

[5] O. Akgul, R. Roberts, M. Namara, D. Levin, and M. L. Mazurek, "Investigating VPN Ads on YouTube," in *5th Workshop on Technology and Consumer Protection (ConPro'21)*, 2021.

[6] R. Ramesh, L. Evdokimov, D. Xue, and R. Ensafi, "VPNalyzer: Systematic Investigation of the VPN Ecosystem," in *Network and Distributed System Security*, 2022.

[7] M. Ikram, N. Vallina-Rodriguez, S. Seneviratne, M. A. Kaafar, and V. Paxson, "An Analysis of the Privacy and Security Risks of Android VPN Permission-enabled Apps," in *Proceedings of the 2016 Internet Measurement Conference*, 2016.

[8] M. T. Khan, J. DeBlasio, G. M. Voelker, A. C. Snoeren, C. Kanich, and N. Vallina-Rodriguez, "An Empirical Analysis of the Commercial VPN Ecosystem," in *Proceedings of the 2018 Internet Measurement Conference*, 2018.

[9] Namecheap, "Google Trends Reveals Surge in Demand for VPN," https://www.namecheap.com/blog/vpn-surge-in-demand/, 2020.

[10] K. Kasunic, "These 7 Companies Secretly Own Dozens of VPNs," https://www.vpnmentor.com/blog/companies-secretly-own-dozens-vpns/, 2021.

[11] S. Taylor, "Kape Technologies (Formerly Crossrider) Now Owns ExpressVPN, CyberGhost, Private Internet Access, Zenmate, and a Collection of VPN "Review" Websites," https://restoreprivacy.com/kape-technologies-owns-expressvpn-cyberghost-pia-zenmate-vpn-review-sites/, 2021.

[12] S. Taylor, "These VPN "Review" Websites are Actually Owned by VPNs," https://restoreprivacy.com/vpn-review-websites-owned-by-vpns/, 2021.

[13] J. Youngren, "Hidden VPN owners unveiled: 104 VPN products run by just 24 companies," https://vpnpro.com/blog/hidden-vpn-owners-unveiled-97-vpns-23-companies/, 2022.

[14] L. Richardson, "Beautiful soup 4.9.0 documentation," https://www.crummy.com/software/BeautifulSoup/bs4/doc/.

[15] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Transactions on Information Systems (TOIS)*, 2010.

[16] M. G. Kendall, "Rank correlation methods." 1948.

[17] Y. Grauer and S. Blair, "Security and Privacy of VPNs Running on Windows 10," https://digital-lab-wp.consumerreports.org/wp-content/uploads/2021/12/VPN-White-Paper.pdf, 2021.

[18] Y. Grauer, "Mullvad, IVPN, and Mozilla VPN Top Consumer Reports' VPN Testing," https://www.consumerreports.org/vpn-services/mullvad-ivpn-mozilla-vpn-top-consumer-reports-vpn-testing-a9588707317/, 2021.