Poster: A First Look Into Distributed VPNs

Yunming Xiao⋄, Matteo Varvello⋆, Aleksandar Kuzmanovic⋄
⋄ Northwestern University, ⋆ Nokia Bell Labs
yunming.xiao@u.northwestern.edu,matteo.varvello@nokia.com,akuzma@northwestern.edu

1 INTRODUCTION

Ookla [9] reported that average Internet speeds in American homes grew 20x in the last 10 years, from 8 Mbps in 2010 to 180 Mbps in 2021, and a similar growing trend is observable worldwide [8]. Recent studies [11, 13] suggest that the added cost for faster Internet speeds — e.g., $50 monthly to boost from 200 to 300 Mbps with Comcast Xfinity [2] — is not worth to most residential users, which consume only 5% of their available bandwidth. Several systems were launched which allow users to monetize such *spare* bandwidth, e.g., distributed proxy or VPN (Virtual Private Network) services [4, 5, 7, 10].

While such systems are attracting a considerable number of users, both as clients (buyers) and providers (sellers), little to nothing is known about such *bandwidth marketplaces*. We have identified *distributed VPN* (dVPNs) as the most concrete examples of such marketplaces because, in essence, a dVPN is a system which allows users to sell their spare bandwidth. The contribution of this paper is an investigation of the dVPN ecosystem. We run *active* and *passive* measurements to characterize their footprint and performance.

The analysis of up to six months of dVPNs data shows that the three major dVPNs (Mysterium [4], Sentinel [7], Tachyon [10]) compose together a fast-growing network footprint of thousands of nodes (sellers) located all over the world. These nodes offer download speeds comparable with ProtonVPN [6], a popular centralized VPN. Location-wise, the US is the most attractive market, with the majority of the traffic being destined to services located in the US.

2 BACKGROUND

DVPN users have two roles, either concurrently or disjointly: *node* and *client*. A user acts as a node when it forwards traffic on behalf of other users. A user acts as a client when its traffic is tunneled via a dVPN node. Hola [1] and VPN Gate [12] are early attempts to dVPNs but they are not useful for our study since their users are not allowed to charge for their bandwidth, thus not making them good approximations of actual bandwidth marketplaces. On the contrary, more recent dVPNs like Mysterium [4], Sentinel [7], Tachyon [10] are appropriate candidates since they allow users to receive some cryptocurrency in exchange of their bandwidth. These dVPNs were selected for our study also because they offer stable client and node implementations.

3 DATA COLLECTION AND ANALYSIS

Data Collection – We *actively* and *passively* collect data from dVPNs. Active experiments consist of automating a dVPN client to discover and connect to available nodes. This allows to discover a dVPN footprint, pricing, and performance. Passive experiments consist of contributing bandwidth to such dVPNs by running several nodes. This is useful to characterize how much and which traffic a typical dVPN node carries.

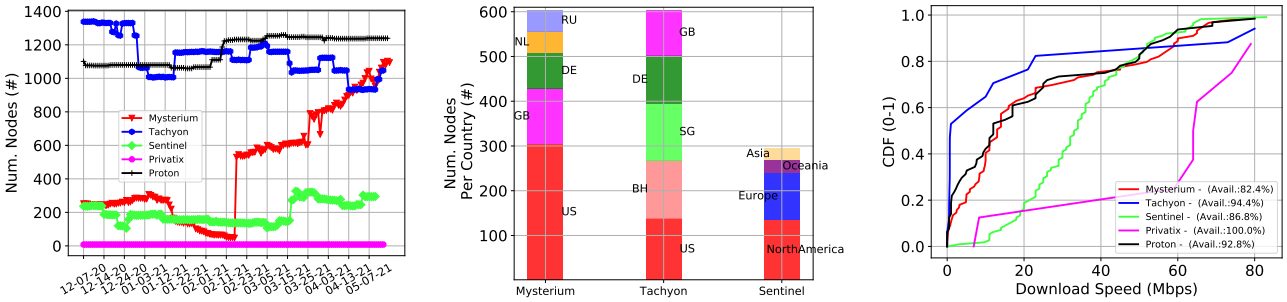
For active experiments, our testbed consists of a powerful Android device with abundant bandwidth controlled by a Raspberry Pi 4. The Android device runs each dVPN app, while the Raspberry Pi realizes the automation. We iterate through the GUI of each dVPN to gather general information and periodically connect to the nodes for speed tests.

For passive experiments, we have deployed 10 machines running nodes for mentioned dVPNs across 4 countries (US, UK, Italy, and China). Our passive measurements last for 3 months (February to April 2021) and account for ∼16 TB of traffic. The IRB at our institution has determined that our work is not considered human research.

Footprint and Performance – Figure 1(a) shows, for each dVPN, the evolution over the last six months (December 2020 – May 2021) of the *total* number of nodes advertised by each dVPN. The figure is further enhanced with data collected from ProtonVPN [6], a popular centralized VPN, given a basic account ($5 per month). Figure 1(a) shows that, originally, only Tachyon had a footprint comparable with ProtonVPN, *i.e.,* in the order of one thousand nodes. However, Tachyon has lost 36% of its nodes over time while Mysterium's node count has been steadily increasing in the last three months, and it is now the largest dVPN with 1,100 nodes.

Next, we report on *where* dVPN nodes are located. Each stacked barplot in Figure 1(b) shows the top 5 countries per dVPN. The figure shows that, irrespective of the dVPN, the US (NorthAmerica for Sentinel) is currently the country where most nodes are located. Germany (DE) and Great Britan (GB) are two other popular countries among dVPNs.

Finally, we report on the *performance* – in terms of download speed and availability – when using such dVPNs. Figure 1(c) shows the Cumulative Distribution Function (CDF) of the download speed measured each month per dVPN (plus ProtonVPN). The figure shows that only Tachyon is overall



(a) (b) (c)

Figure 1: Footprint and performance characterization of the dVPN ecosystem: (a) Evolution over 6 months of the number of nodes; (b) Histogram of number of nodes per country; (c) CDF of download speed.

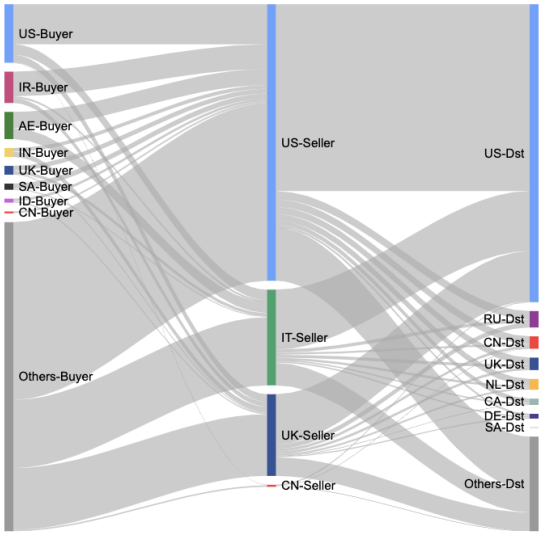


Figure 2: Visualization of dVPN traffic across Mysterium, Sentinel, and Tachyon. Buyer's locations are shown on the left, our machines where dVPN nodes are run in the center, and traffic destinations on the right.

slower than ProtonVPN. Mysterium has comparable performance with ProtonVPN while both Sentinel and Privatix significantly improve bandwidth, by up to 3x and 6x. The legend of Figure 1(c) also reports the overall availability of each dVPN which is, on average, comparable with ProtonVPN.

Traffic Characterization – Between February and April 2021, our nodes have served ~632 thousand dVPN sessions, ~623 million TCP/UDP flows, accounting for about 16 TB of data. Download traffic is the highest contributor, about 10x the amount of upload traffic. Figure 2 visualizes from where dVPN traffic originates and is destined to, using Maxmind [3] to map `ip_src` and `ip_dst` at the country level. The middle of the plot shows the 10 machines – distributed between US, Italy, UK, and China – which were used for passive data collection. The figure aggregates data across the three dVPNs since no statistically meaningful difference was observed. The figure shows that the US has the most buyers, followed by Iran (IR) and United Arab Emirates (AE). The US is also the most popular destination regardless of which node (middle of the plot) is used, accounting for over half of the traffic. Russia (RU) is the second most popular destination, followed by China (CN), UK, and Netherlands (NL).

4 FUTURE WORK

We plan to continue exploring the dVPN ecosystem with focus on the traffic *characteristics*, e.g., presence or lack of harmful traffic, to help asses the *risk* associated with running a DVPN node. We are further interested in understanding the economics between buyers and sellers. We will study the value of spare bandwidth in today's bandwidth marketplace and further explore the opportunities of optimizing the buyers' costs and sellers' income.

REFERENCES

[1] Hola: Get access to worldwide content. https://hola.org/.
[2] Internet Deals, Plans, and Pricing | Xfinity. https://www.xfinity.com/learn/internet-service/deals.
[3] MaxMind: IP Geolocation and Online Fraud Prevention. https://www.maxmind.com/.
[4] Mysterium network: Censorship free Internet for all. https://mysterium.network/.
[5] Privatix: Next-gen VPN. https://privatix.com/.
[6] ProtonVPN. https://protonvpn.com/.
[7] Sentinel: Secure yourself today with Sentinel. sentinel.co.
[8] Speedtest Global Index. https://www.speedtest.net/global-index.
[9] Speedtest Global Index - United States. https://www.speedtest.net/global-index/united-states.
[10] Tachyon VPN. https://tachyon.eco/.
[11] The Truth About Faster Internet: It's Not Worth It, The Wall Street Journal, 2019. https://www.wsj.com/graphics/faster-internet-not-worth-it/.
[12] VPN Gate - Public VPN Relay Servers. https://www.vpngate.net/en/.
[13] F. Bronzino, P. Schmitt, S. Ayoubi, G. Martins, R. Teixeira, and N. Feamster. Inferring streaming video quality from encrypted traffic: Practical models and deployment experience. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(3):1–25, 2019.