Barcelona School of Informatics

Bachelor’s degree in Informatics Engineering

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Bachelor’s Thesis

Is Your VPN Trustworthy

Author

Ismael de la Garcia Cobos

Advisors

Pere Barlet Ros

Ismael Castell

Q1 2020/2021

Abstract

Nowadays people are more and more concerned about their privacy and safety on the internet. Also, being public as it is, some content is limited based on geolocation. On the other hand, remote work is also becoming a matter of fact. These things have empowered the use of Virtual Private Networks. However, it is well known that VPNs, if malicious, can steal all your personal information. In this paper we want to build a tool that can analyze and detect any malicious traffic in your VPN.

Resumen

Cada día la privacidad y la seguridad en internet es un tema que preocupa a más personas. Además, y a pesar de ser público, hay contenido que se limita dependiendo de la localización. Por otro lado, el trabajo en remoto está cada vez más a la orden del día. Todo esto ha hecho que crezca el uso de las Virtual Private Networks. Es bien sabido que una VPN con malas intenciones puede llegar a robar toda tu información personal y privarte de la privacidad. En este trabajo se pretende crear una herramienta que permita detectar trafico malicioso en un cliente VPN.

Resum

Cada dia la privacitat i la seguretat a internet és un tema que preocupa a més persones. A més, i tot i ser públic, hi ha contingut que es limita depenent de la localització. D'altra banda, el treball en remot està cada vegada més a l'ordre del dia. Tot això ha fet que creixi l'ús de les Virtual Private Networks. És ben sabut que una VPN amb males intencions pot arribar a robar tota la teva informació personal i privar-te de la privacitat. En aquest treball es pretén crear una eina que permeti detectar trànsit maliciós en un client VPN.

Revision history and approval record

Version	Date	Description
0	28/09/2020	Document creation

Table 1: Record of modifications made to this document

List of Figures

List of Tables

Table of contents

Project scope

A new way of working is approaching due to the recent changes in the world. Remote work has become a necessity in countries which were not ready for it. On the other hand, our privacy is threatened every day to a point most people can not understand. That is why many companies take advantage of this situation to make money out of it. It is widely spread that VPNs, specially the free versions, offer less than paid services. Many people state that free services may also sell the users privacy online to third party companies while saying they offer increased privacy and security.

Given this situation, I found the need of a tool that can protect users from being betrayed by their VPN providers on a daily basis. Honesty should be the key to any agreement, even when using free software. Therefore, in this paper I want to explain a tool that can detect malicious traffic in a VPN client, specially the kind of traffic that may compromise the user's privacy on purpose.

Therefore, this work has the following objectives:

- Learn about Online Privacy and how VPNs can improve it.
- Find common mistakes in VPN clients that can compromise the user's online privacy.
- Learn tracking methods VPN clients can use.
- Detect tracking methods related to traffic analysis and online behaviour.
- Create a tool that can detect VPNs using this tracking methods.

1.1 Requirements and specifications

The requirements of the project come from the need of portability of this tool, for users of VPNs may use very different environments.

- The tool should work in the main OS (Windows, Linux, Mac).
- The tool should work with as much VPN clients as possible.
- The tool should be user-friendly so that all sorts of people can benefit from it.

1.2 Actors

In this section I will list the stakeholders of this thesis, be them people who will benefit, be affected partake of it.

- Developer: The project will have a single developer, Ismael de la Gracia Cobos, as part of his bachelor's dregree. He will research, develop, test and document the project.
- Advisors: They will guide the developer in the project. The main advisors are Pere Barlet and Ismael Castell, a teacher in charge of many projects related to internet privacy and a doctorate student on internet privacy respectively.
- VPN Providers: VPN providers will definitely be affected by this project, for their services may become compromised after the development of the proposed tool. Any VPN provider conducting privacy-abusing practices may be exposed by this tool. However, all those who do not perform wrong practices will receive the support of a third party tool in their claims for privacy and security.
- Advertising companies: Every VPN that implements tracking in their clients later sells that information to an advertising company. The development of a tool that detects this interaction will surely punish this type of relations.
- IT Departments: Most companies now a days have an IT department that takes care of the installation of the necessary infrastructure for the company to work. These departments may benefit out this tool in auditing their systems for security, for they will be able to know if they are using real private networks.
- Tech-savvy users: This kind of users usually care about their privacy when connected to the internet. Most at times, they use VPNs to access restricted content or to navigate the internet privately. They may also wonder to what extent they are doing so. This tool will surely be interesting to them.

1.3 Obstacles and risks

Throughout the project, we may encounter unexpected obstacles and risks that we will need to affront. Here we put some examples:

- Implementation errors: Sadly, it is very common that developers make mistakes, for we are all humans. Debugging the type of tool we want to create may be hard, for it works at a very low level and the developer, Ismael de la Gracia, has no previous experience in this field.
- Delays: No room for distraction. The project has to progress carefully to arrive on time to the deadline. If we delay on something, we can cause the delay of the other projects.
- Misconceptions: Since the developer has no experience in this field and all the knowledge has just been acquired by self investigation, there might be some misconceptions that may lead to poor decisions. Therefore, it is important to invastigate properly and to trust in the guidence of the advisors.
- Bad documentation: Trying to understand how a project is done is a waste of time if it does not have good documentation. It is important to document every progress and change we make to the project.

Definition and Justification

2.1 Definition

A VPN (Virtual Private Network) extends a private network across a public network and enables users to send and receive data across shared or public networks as if their computing devices were directly connected to the private network [16]. For security, the private network connection may be established using an encrypted layered , and users may be required to pass various authentication methods to gain access to the VPN. Nowadays, OpenVPN is the one of the most used protocols available.

VPNs are often considered the key to safety and privacy while browsing the internet. However, there are some important things that we have to take into consideration when choosing a VPN if we want to enjoy safety and privacy.

2.1.1 Supported devices

VPNs are supported in many OS. Some of them are Windows, Mac OS X, iOS, Android, and Linux (Ubuntu) and Chrome book. They also work on mobile devices, like Android, IPhone and IPad. Some routers also support VPNs.

2.1.2 Privacy

While most VPN providers claim they do not keep logs, it is true that most companies keep basic connection logs(User IP address and VPN timestamp). It is very important to check the companies privacy policy and the location of their headquarters (in terms of jurisdiction).

It is also important to take into account the encryption standard used. Many VPN clients allow you to choose different standards. While the securer standards may slow down the internet connection, they will make sure your data is safer. You may find more information on encryption in the appendix, section 4.1.

Typically, your ISP (Internet Service Provider) would have to pay special attention to your activity in order to determine whether you are using a VPN or not. However, if well protected, they will not be able to learn the nature of your activity. They are only able to see you are connected to a certain endpoint and the bandwidth you are using. If they want to learn more about you, they may request logs to your VPN provider (they probably need a court order to do that).

Strictly speaking, using a personal VPN does not prevent others from tracking you fully. Some tracking techniques can go through a VPNs layer of privacy and still identify you while browsing the internet. Some of this practices are simple, some others are more complicated. See appendix for more detailed information, on section 4.2.

2.1.3 Security

Notice that using a VPN does not protect you from malicious software. While some VPN providers offer additional protection, most do not. Viruses, spyware and malware can still harm you when you are using a VPN.

2.2 Justification

Virtual Private Networks (VPNs for short) are widely used in today's world. Particular users normally use them to increase their privacy while using the internet. Others want to access content that is blocked in their country or location. Companies often use VPNs to access their resources remotely in a way these resources can not be stolen or seen by other users on the internet. VPNs are often linked with security and privacy. On the other hand, many people suspect that some VPN providers are selling their online activity to advertisement companies, but can not prove it true. Huge stirs have been raised when researchers prove a certain provider is violating the users privacy, but most of the times they have to prove it by analyzing the code of the VPN client [8]. In this thesis, we want to create a tool that can distinguish the legitimate and malicious traffic a VPN may generate. We identify as malicious traffic all the traffic or activity a VPN may generate that may compromise the users anonymity on the internet. Therefore, we do not want to create an antivirus software, but an anti-tracking tool.

Virtual Private Networks pretend to fight against privacy-abusive practices, or at least that is what they sell to their customers. When a user decides to use the services of a VPN provider, most of the times they have to read carefully their privacy policy and place their trust on the company that will provide them these services.

Therefore, and given that there is no similar tool to the one described in this document, I found the need of an automated tool that helps detect illegitimate traffic a VPN client might generate.

With this tool we want to focus on the following tracking techniques. We also contemplated DNS/IPv6 leaks as a dangerous practice that can lead to the users's privacy being compromised, but since this is not result of a malicious practice but rather comes from wrong programming, the tool does not have to catch this case.

- JavaScript Injection.
- TLS Interception.
- Traffic redirection.

Methodology

The development of the project will be done using the Agile methodology, specifically Scrum [12]. Scrum is an iterative and incremental methodology, in such a way that a number of short iterations (sprints) of between 1 and 3 weeks are defined, and in each iteration value is added to the final product. This methodology divides the team into different roles. However, since there is only one developer in this project, he will carry all the tasks. In a Scrum project, a board must initially be filled (backlog) with the user stories punctuated according to the difficulty of each, and each sprint is assigned some of these user stories that will be divided between the team of developers, and must be completed in the end of the sprint. To make it easier to develop following this Scrum methodology, several tools will be used.

3.1 Tools

3.1.1 Git

Git [5] is a free and open source version control system. It is widely used in software development in cooperative projects to perform version management, through repositories and branches.

3.1.2 Github

GitHub [1] is a free software Git repository manager that also has other features such as error control, and continuous integration.

3.1.3 Git-flow

The Git-flow methodology [4] consists of a always stable Master Master branch, a Development branch, and Release and Feature branches. The git-flow methodology proposes that for each

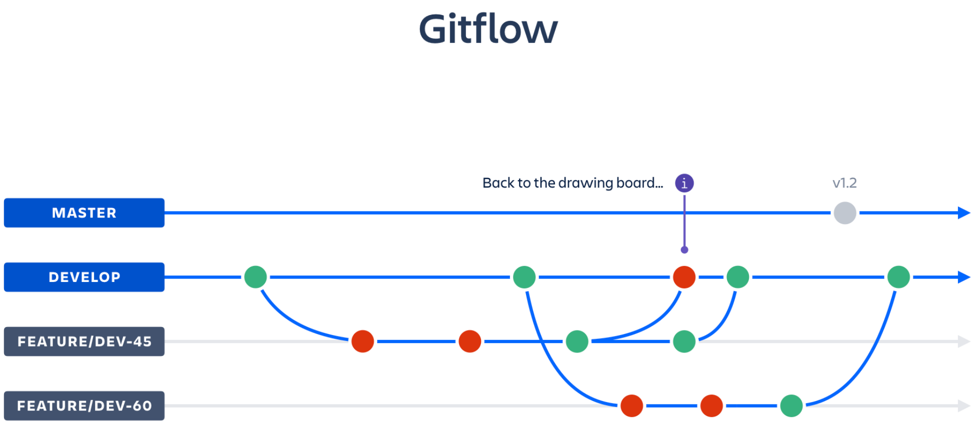


Figure 3.1: Git-flow example. Source: Atlassian.com

user history that is implemented, a new branch of Feature is created with origin in the branch of Development, and that once this user history was complete the header is developed in Development. Figure 3.1 shows an example.

3.1.4 Taiga

Taiga [10] is a free and open source project management system. It has features to create and rate user stories, to manage the backlog, and to track issues.

3.1.5 Slack

”Connect your conversations with the tools and services that you use to get the job done. With over 2,000 apps and a robust API, the Slack [15] platform team works with partners and developers globally to build apps and integrations that streamline your work, automate mundane tasks and bring context into your conversations in Slack”. We will use this tool to comunicate with the rest of the students and with the advisors.

Appendix

4.1 Data encryption

Essentially, data encryption provides users with a secure VPN tunnel whereby data is protected. However, there are different levels of encryption as well. By default OpenVPN includes AES-128 [2], SHA-1 [13], and RSA-2048 [11], while maximum encryption includes AES-256 [13], SHA-256 [14], and RSA-4096 [11]. An encryption key tells your computer how to decrypt or encrypt data.

Data authentication is part of the encryption process and refers to the message authentication algorithm with which user data is authenticated. This is used to protect users from active security attacks.

4.2 Tracking Techniques

4.2.1 Non-VPN Related

Tracking cookies

While using a VPN can hide your IP address from the internet, cookies [7] can still track your activity. Tracking cookies are not blocked by your VPN when browsing the internet, and they are not removed from your system afterwards either. However, your web browser can do that for you. Every web browser can block third-party cookies and remove all of them after the browsing session has finished (when you exit the browser), and if yours does not, you should think about changing your web browser. Sometimes this is as simple as browsing in incognito mode [6].

A VPN, however, can help in the sense that it will hide from cookies your real IP address, since the VPN server is the one connecting to the internet. Therefore, IP based identification will not work on you. But all the other information will be there: browsing history, queries performed, preferences...

Referrer URLs

A referrer URL is the web address of the previous website where you clicked a link to get to the current website. These can be used for several reasons, and recording your browser history is one of them. They are often used alongside with cookies, but these are not required in order to use referrer URLs. If someone clicks on a link to `example.org` at `example.com/links.htm`, then `example.org's` visitor log will show `example.com/links.htm` as his referral URL [9].

Some web browsers have an option not to send referral URL information to websites. Generally speaking, VPNs do not do that. Users may also do this manually by copying the link address and pasting it into the browser's address bar.

Web beacons

A web beacon is a technique used on web pages and email to unobtrusively (usually invisibly) allow checking that a user has accessed some content [17]. The first web beacons were small digital image files that were embedded in a web page or email. The image could be as small as a single pixel, and could be of the same color as the background, or completely transparent (thus the name "tracking pixel"). When a user opens the page or email where such an image was embedded, they might not see the image, but their web browser or email reader would automatically download the image, requiring the user's computer to send a request to the host company's server, where the source image was stored. This request would provide identifying information about the computer, allowing the host to keep track of the user.

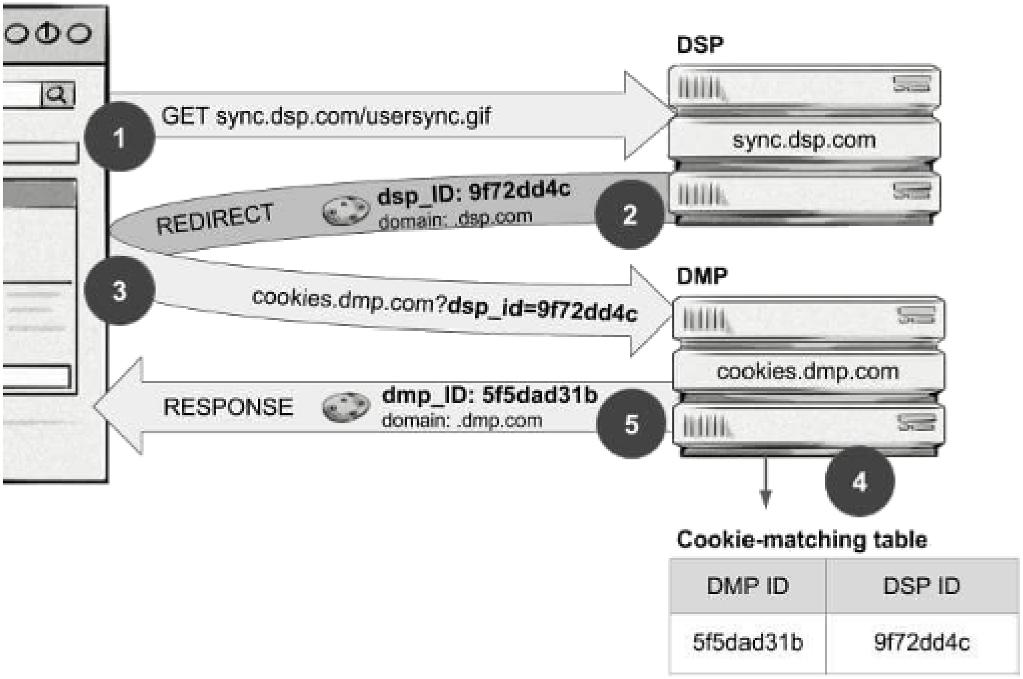
The identifying information provided by the user's computer typically includes its IP address, the time the request was made, the type of web browser or email reader that made the request, and the existence of cookies previously sent by the host server. The host server can store all of this information, and associate it with a session identifier or tracking token that uniquely marks the interaction.

Browser fingerprinting

A device fingerprint, machine fingerprint or browser fingerprint is information collected about a remote computing device for the purpose of identification. Fingerprints can be used to fully or partially identify individual users or devices even when cookies are turned off [3].

The attributes of the fingerprint depend on the web application, for each of them may use a different set of attributes. For reference, `amiunique.org` uses the following data to create their fingerprint:

- the User agent header
- the Accept header
- the Connection header
- the Encoding header
- the Language header
- the Upgrade Insecure Requests header
- the Referer header
- the Cache-Control header
- the BuildId of the browser
- the list of plugins
- the platform
- the cookies preferences (allowed or not)
- the Do Not Track preferences (yes, no or not communicated)
- the timezone
- the screen resolution and its color depth
- the use of local storage
- the use of session storage
- a picture rendered with the HTML Canvas element
- a picture rendered with WebGL
- Supported Audio formats
- Supported Video formats
- the presence of AdBlock



DMP ID	DSP ID
5f5dad31b	9f72dd4c

Figure 4.1: Image syncing example. Source: Google Images

- the list of fonts

Nor VPNs, nor most ad-blockers can protect users from this type of tracking. Some ad-blockers are working on blocking this type tracking by removing the HTML canvas element from a web page when certain conditions are met, for this is usually used for tracking purposes.

Cookie syncing

The biggest problem cookies face is that they can only be read on the domain that created them. This means that AdTech companies can't read cookies created by other AdTech platforms or by the website itself, essentially limiting their effectiveness for advertising purposes on other websites. This is faced by creating a unique identifier (UID) for each user and sharing it with other companies, like a translation table that gathers each companies ID of that user and translates it to a UID.

Supercookies

Supercookies are not stored in your computer. Instead, an ISP inserts a piece of information unique to a user's connection into the HTTP header. They are injected at the network level as

Unique Identifier Headers (UIDH). This means clearing your browser data will not delete that cookie. Blockers can't block it either.

A zombie cookie remains intact as it hides outside of your browser's regular cookie storage. Zombie cookies target local storage, HTML5 storage, RGB color code values, Silverlight storage, and more. That's why they're known as zombie cookies. An advertiser must only find an existing cookie in one of those locations to resurrect the rest.

If an ISP decides to track you with supercookies, there is not really much you can do, except encrypting your traffic (so that at least, they do not know what you are doing). This can be done by using HTTPS-only websites, and by using a VPN. However, the last if the safer option.

4.2.2 VPN Related

To be expanded.

Bibliography

[1] Inc 2020 GitHub. Build software better, together. GitHub. Feb. 8, 2008. URL: https://github.com (visited on 09/29/2020).

[2] Advanced Encryption Standard. In: Wikipedia. Page Version ID: 980596722. Sept. 27, 2020. URL: https://en.wikipedia.org/w/index.php?title=Advanced_Encryption_Standard&oldid=980596722 (visited on 09/29/2020).

[3] Device fingerprint. In: Wikipedia. Page Version ID: 973166516. Aug. 15, 2020. URL: https://en.wikipedia.org/w/index.php?title=Device_fingerprint&oldid=973166516 (visited on 09/29/2020).

[4] Vincent Driessen. Using git-flow to automate your git branching workflow. Aug. 19, 2010. URL: https://jeffkreeftmeijer.com/git-flow/ (visited on 09/29/2020).

[5] git-scm.org. Git. Git. Apr. 10, 2005. URL: https://git-scm.com/ (visited on 09/29/2020).

[6] How are Cookies Affected by a VPN? VPNPros. Section: Blog. Oct. 2, 2018. URL: https://vpnpros.com/blog/how-cookies-affected-by-vpn/ (visited on 09/29/2020).

[7] HTTP cookie. In: Wikipedia. Page Version ID: 977365598. Sept. 8, 2020. URL: https://en.wikipedia.org/w/index.php?title=HTTP_cookie&oldid=977365598 (visited on 09/29/2020).

[8] Muhammad Ikram et al. "An Analysis of the Privacy and Security Risks of Android VPN Permission-enabled Apps". In: Proceedings of the 2016 ACM on Internet Measurement Conference - IMC '16. the 2016 ACM. Santa Monica, California, USA: ACM Press, 2016, pp. 349–364. ISBN: 978-1-4503-4526-2. DOI: 10.1145/2987443.2987471. URL: http://dl.acm.org/citation.cfm?doid=2987443.2987471 (visited on 09/29/2020).

[9] William Jensen. What Is a Referral URL? Techwalla. 2020. URL: https://www.techwalla.com/articles/what-is-a-referral-url (visited on 09/29/2020).

[10] Taiga Agile LLC. Taiga.io. Taiga - Love your projects. Sept. 12, 2016. URL: https://taiga.io/ (visited on 09/29/2020).

[11] RSA (cryptosystem. In: Wikipedia. URL: https://en.wikipedia.org/wiki/RSA_(cryptosystem (visited on 09/29/2020).

[12] Scrum.org. What is Scrum? Scrum.org. Mar. 30, 2010. URL: https://www.scrum.org/resources/what-is-scrum (visited on 09/29/2020).

[13] SHA-1. In: Wikipedia. Page Version ID: 979702000. Sept. 22, 2020. URL: https://en.wikipedia.org/w/index.php?title=SHA-1&oldid=979702000 (visited on 09/29/2020).

[14] SHA-2. In: Wikipedia. Page Version ID: 977780884. Sept. 10, 2020. URL: https://en.wikipedia.org/w/index.php?title=SHA-2&oldid=977780884 (visited on 09/29/2020).

[15] Slack. El motor de tu trabajo. Slack. 2020. URL: https://slack.com/intl/es-es/ (visited on 09/29/2020).

[16] Virtual private network. In: Wikipedia. Page Version ID: 979378541. Sept. 20, 2020. URL: https://en.wikipedia.org/w/index.php?title=Virtual_private_network&oldid=979378541 (visited on 09/29/2020).

[17] Web beacon. In: Wikipedia. Page Version ID: 973560171. Aug. 17, 2020. URL: https://en.wikipedia.org/w/index.php?title=Web_beacon&oldid=973560171 (visited on 09/29/2020).