Bypassing Cloudflare's reverse proxy: a case study

Contornando o proxy reverso do Cloudflare: um estudo de caso

DOI:10.34117/bjdv8n4-298

Recebimento dos originais: 21/02/2022
Aceitação para publicação: 31/03/2022

Leandro de Souza Oliveira
Expert in Software Engineering, Computer Forensics and Digital Forensics
Institution :Instituto de Criminalística, Polícia Civil do Distrito Federal. Brasília/DF Brasil
Address: SPO Conjunto A, Lote 23, Complexo da PCDF Parque da Cidade Sudoeste Brasília DF
E-mail:leosol@gmail.com

João Paulo Claudino de Sousa
M.Sc. Electrical Engineering, Computer Forensics and IT Security
Institution: Instituto de Criminalística, Polícia Civil do Distrito Federal. Brasília/DF Brasil
Address:SPO Conjunto A, Lote 23, Complexo da PCDF Parque da Cidade Sudoeste Brasília DF
E-mail:jpclaudino@gmail.com

João Vitor Assis Ribeiro
Bachelor in Computer Science
Institution: Instituto de Criminalística, Polícia Civil do Distrito Federal. Brasília/DF Brasil
Address: SPO Conjunto A, Lote 23, Complexo da PCDF Parque da Cidade Sudoeste Brasília DF
E-mail:joao.ribeiro@pcdf.df.gov.br

ABSTRACT

Reverse proxy is a functionality provided by companies such as Cloudflare and its designed to protect virtual assets on the internet by acting as a middleman between end users and an origin server. While working on a law enforcement case, we performed OSINT research, designed and deployed a tool that allowed us to reach an original website's IP address associated with criminal activity, thus achieving partial bypass of Cloudflare's reverse proxy protection.

Keywords: computer networks, Cloudflare, reverse proxy bypass.

RESUMO

Proxy reverso é uma funcionalidade fornecida por empresas como Cloudflare e é projetado para proteger ativos virtuais na internet, atuando como servidor intermediário entre usuários finais e um servidor de origem. Ao trabalhar em um caso de investigação policial, nós realizamos buscas em bases abertas, implementamos e implantamos uma ferramenta que nos permitiu chegar ao endereço IP original de uma página da web

associada a atividade criminosa, conseguindo, assim, burlar parcialmente a proteção de proxy reverso fornecido pela empresa Cloudflare.

Palavras-chave: redes de computadores, Cloudflare, burlamento de proxy reverso.

1 INTRODUCTION

In today's world, there is an increasing effort, by businesses, corporations and even malicious agents, towards safeguarding digital assets in the connected, digital world. From legitimate to unlawful actors, they all wish to take advantage of the Web's visibility to showcase their services and products while still retaining (some) anonymity.

From fake auction sites to illegal drug trafficking, some criminals put a web page in the World Wide Web at plain sight, serving either as a vitrine or mediator of criminal activity, while also employing security practices with the goal of law avoidance.

One of these practices is using CDNs - Content Delivery Networks - and DDoS - Distributed Denial of Service - mitigation services, in which, among other functionalities, serve as a middleman between an ASP - Application Service Provider and the end user, a service which is commonly referred to as reverse proxy.

One of the most prominent providers of such functionality is Cloudflare [1], providing security services such as Web Application Firewall, and DNS - Domain Name System, acting as a reverse proxy between the application provider and the end-user. As such, searches done on search engines, e.g., Google and in DNS-lookup services typically point to Cloudflare servers instead of the website's service provider.

The desired - or undesired, depending of the angle - consequence of using said service is that it is often inconvenient for law enforcement and prosecution to seize access to a website if the original Application Service Provider's IP is unknown and cooperation from the reverse proxy provider is either denied, unsatisfactory or not provided in a timely manner.

That said, there have been a few cases [3] in Brazil where law enforcement agents were forced to resort to procedures such as assessment and exploitation of vulnerabilities of a web page hosting and/or facilitating criminal activities, which led to the indictment and prosecution of actors, a process - relatively new in Brazilian Criminal Justice System - known as Legal Hacking.

In such law enforcement scenarios, or in usual pentest situations, the feature of reverse proxy might prove a challenge. For dealing with such challenge, we, while

working on a case of Legal Hacking of a web page used in illegal drug trafficking, performed OSINT research and designed a simple tool that allowed us to reach the website's ASP IP address, thus partially circumventing Cloudflare's reverse proxy functionality.

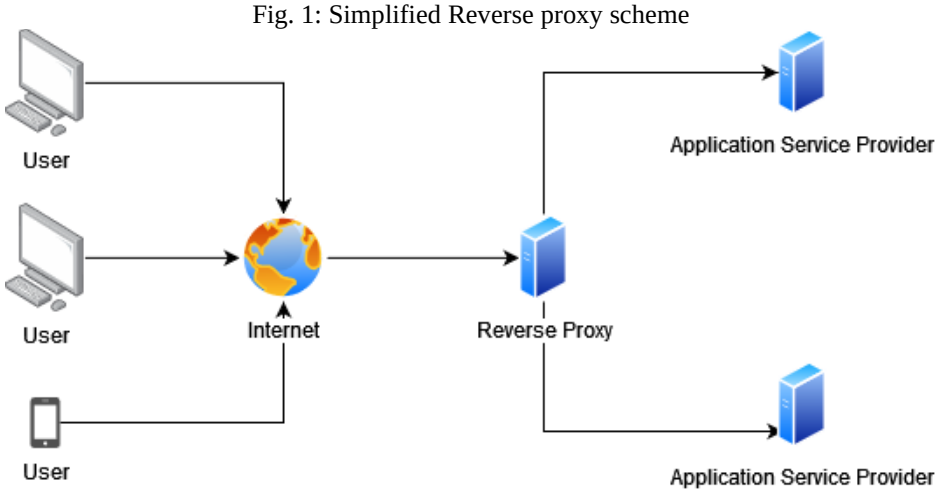
2 CLOUDFLARE'S REVERSE PROXY - A BRIEF REVIEW

A reverse proxy is a server that acts as a middleman between an Application Service Provider and its end users wishing to access a web page or service. It usually provides two main functionalities: 1. Domain name resolving and 2. Forwarding client requests - while performing a range of security filters, input handling and other functionalities - to the original server [4].

There are many reasons for why one would use such service, and a non exhaustive list is presented as follows:

- To detect and effectively protect against cyberattacks such as DDoS, intrusion, data exfiltration, ransomware, and others
- Perform request processing and optimization (web acceleration), caching, and additional tasks such as SSL encryption and overall traffic optimization [5].
- To perform load balancing of web pages with huge incoming traffic unfeasible to be handled on a single origin server, providing fail safety and redundancy;
- Implement anonymity by hiding the identity of the original ASP server owner, including leasing, licensing and certificate information, limiting the possibilities of users tracing the original web page owner or legal representative.

Fig. 1: Simplified Reverse proxy scheme



3 RESERVE PROXY BYPASS

The central security premise is that a website configured with Cloudflare is only accessible by end users through Cloudflare itself, but that is not always is the case, as origin servers are not always configured properly, so the exposure and/or vulnerability lies not in the service itself but in its misconfiguration.

Given that, it is arguably that this misconfiguration is the main target of attacks aimed at getting behind Cloudflare's protection. That is, reaching the origin server IP address. One of the most common scenarios is that Cloudflare's servers are whitelisted but incoming outside traffic is not prevented, thus the reverse proxy service only addresses the name-resolving aspect of the functionality for typical users.

Another possibility is the presence of improper subdomain configuration where they point to the original IP address, while only the main domains are set up with Cloudflare. Baloch [4] also proposes another method of reaching the origin server IP address by tracking, for example, registration mail header information, as Cloudflare does not proxy SMTP traffic [6].

Also, another common method of reaching the origin server IP address is to consult DNS resolvers that keep a watch list of websites that had previous DNS history before transitioning to using Cloudflare's name resolving [4]. Today, there are many online services that log and store internet-wide scan data of web page information and DNS record history. For example, Security Trails [7] Shodan [8], Censys [9] are a few.

Even though a properly configured ASP can be set up to only accept incoming traffic from Cloudflare [10], obtaining the original website IP address might be sufficient for law enforcement and prosecution to reach the perpetrator by requesting the ASP for information or issuing a search and seizure warrant given the origin IP addresses.

In summary, the process of obtaining the IP address of a web site or service utilizing Cloudflare (reaching the origin server IP address) can be laid out as follows:

- Perform a typical pentest approach, looking for possible misconfiguration of the reverse proxy, references or information that might lead to the origin server(s) IP address(es);
- Leverage the fact that services like Cloudflare typically only handle HTTP/HTTPS traffic, and might leak the origin server's IP address, such as email headers (SMTP traffic). This can be exploited if the website performs some sort of automatic mailing service;

- Perform OSINT research of DNS history in order to possibly find out previous IP addresses or AS networks associated with the target web page or service, while seeking for a match.

4 RELATED TOOLS

CloudFail [11] is a recon tool designed for gathering information about a host protected by Cloudflare. Leveraging DNS misconfiguration and records found on the - now defunct crimeflare.com database. Similarly, CloudFlair [12] is a tool with similar goals, differing in the lookup database used to find exposed IPv4 hosts, using, for that, the Censys API [9].

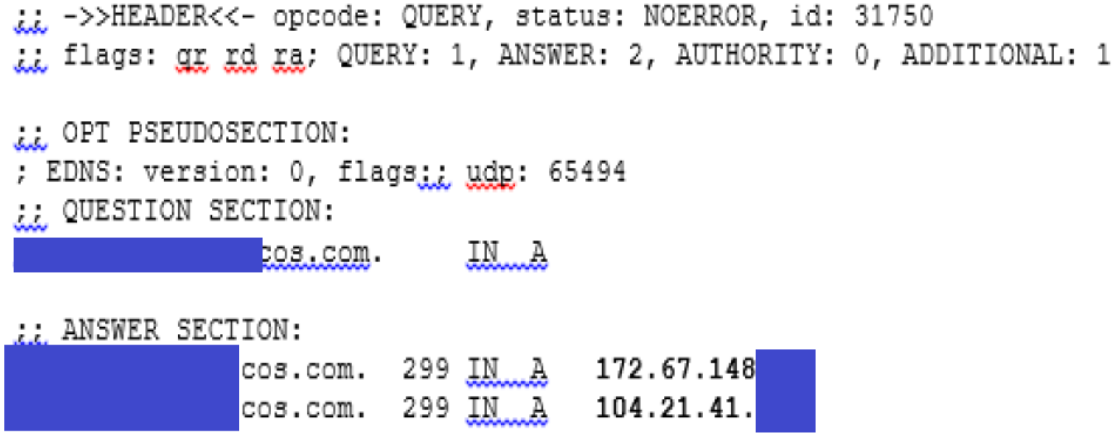
5 LEGAL HACKING - A CASE STUDY

Backed up by a search warrant and legal support, it was presented a case of Legal Hacking of a website hosting propaganda and contact information for illegal drug trafficking activity, while the goal was the adoption of a:

> "*...Procedure of vulnerability assessment of the website (...) remaining authorized while, in the eventual case of the discovery of a security flaw, the seize/extraction of files, documents and emails related to illegal drug trafficking activity linked to the website that may possibly identify and locate its administrators and eventual users associated with illegal drug trafficking...*"

The website consisted of a WordPress 5.5.5 page, with almost no user interaction except a contact link that directed to a WhatsApp web chat with a potential illegal drug dealer. It employed Cloudflare's protection (reverse proxy), and thus an IP probe at the given domain pointed to Cloudflare's servers:

Fig. 2: Website's IP response



Given this, a thorough OSINT research was performed to find registrar information and DNS record history. We thus found another website which shared WHOIS information with the one in investigation, which in turn led to the hypothesis that both these websites would be hosted in the same ASP.

The next step would be to confirm this hypothesis, by obtaining a match between a page served by Cloudflare and one served by the ASP, while hoping that the original ASP was not configured properly and fully accepted direct requests.

6 OBTAINING CANDIDATE NETWORK RANGES

We then proceeded to enumerate possible network ranges (netblocks) for a scan. We found possible candidate ranges in AS databases such as IP Info [13]. We also included in the scan other common service providers in the IANA's LACNIC range (the agency which manages AS - Autonomous System in Latin America) of ASP providers.

7 AS RANGES SCAN

After AS range enumeration, we proceeded to perform various types of scans in hope of finding the origin's ASP IP address, trying, among other tools, the ones mentioned previously, CloudFlair and CloudFail, both without success. The CloudFlair tool was outdated (the database used for matching page information was defunct), and as for CloudFail, no match was found.

8 CLOUDFEIO

This thus led to us developing a tool called CloudFeio [14] - cloud 'ugly' in Portuguese - to perform the scans and automate the process. It is a python tool aimed at identification and reach of a virtual host's IP address that is behind Cloudflare's reverse proxy functionality.

It performs HTTP and HTTPs requests either with SNI - Server Name Identification or without it - by setting the host header - with the presumed server name or identifier string. It does so by probing a provided discrete list of websites or network masks, in which the attacker has obtained, either through OSINT research or by other means in an attempt to receive a expected answer indicative of reaching the final ("real" or virtual) host IP address.

That is achieved by setting a metric which is incremented whenever an expected string is found on the current target address. Also, the tool saves the certificates of the matching websites.

9 AS RANGES SCAN, AGAIN

We then performed another scan with an initial version of CloudFeio (winch at the time, did not include specific string search), while also including a broader range of candidate netblocks ranges.

As we only knew that the page presented by Cloudflare had HTTPS - thus showing a Cloudflare certificate to the end user - and we did not know even if the origin ASP used HTTPS at all, we looked for three types of web page configuration. Plain HTTP, HTTPS without SNI and HTTPS with SNI (virtual hosts) and arrived at 7 (seven) possible matches:

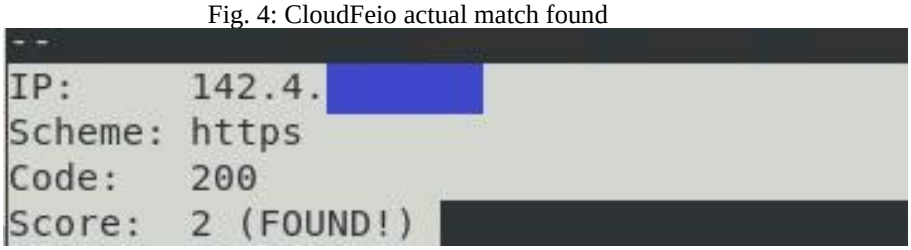
Fig. 3: CloudFeio initial results

```
leandro@debian:~/find-vhost$ cat success-ips.txt
IP:     142.4.[redacted]
Scheme: https
Code:   200
Score:  2 (FOUND!)
--
IP:     144.217.[redacted]
Scheme: https
Code:   200
Score:  2 (FOUND!)
--
IP:     144.217.[redacted]
Scheme: https
Code:   200
Score:  2 (FOUND!)
--
IP:     192.99.[redacted]
Scheme: http
Code:   200
Score:  2 (FOUND!)
--
IP:     192.99.[redacted]
Scheme: https
Code:   200
Score:  2 (FOUND!)
--
IP:     192.99.[redacted]
Scheme: https
Code:   200
Score:  2 (FOUND!)
--
IP:     198.50.[redacted]
Scheme: http
Code:   200
Score:  2 (FOUND!)
leandro@debian:~/find-vhost$
```

We then proceeded to validate the found results. Discovering that WordPress itself was performing some kind of application load balancing, including some content of the target page in other (WordPress) pages. Thus, we adjusted CloudFeio to perform searches of specific expressions present in the target website and (probably) not in these other WordPress pages.

This led to, after further validation, eliminating other IPs and us settling for the first match of the found results (Figure 4):

Fig. 4: CloudFeio actual match found



As that was a HTTPS match, another manual check of the certificate was performed, via a direct browser request. We then found out that it had a different Cloudflare certificate.

Thus, we concluded that the website was set up with Cloudflare's Full Encryption Mode [15], employing one Cloudflare certificate in the front end and another one in the origin:

Fig. 5: Target web page origin ASP certificate information

This led to the attestation of the ASP provider and IP address of the web page hosting criminal activity that was (improperly) configured with Cloudflare's reverse proxy.

10 CONCLUSION

We were able, with OSINT research and the development and deployment of our tool CloudFeio, to achieve a partial Cloudflare proxy bypass, reaching the origin IP of a web page under investigation in a Legal Hacking case, thus aiding law enforcement agents in reaching out for actors possibly involved in criminal activity in the World Wide Web.

We were also able to gain a better understanding of TLS authentication with SNI in web pages set up with reserve proxy, by the appropriate setting of SNI in web page request in automated web page scan (crawling), while also correlating it with the mentioned reverse proxy mechanism.

REFERENCES

[1] John E. Dunn for Forbes (2020). How Cloudflare Became The Most Important Internet Company Nobody Has Heard Of. *https://www.forbes.com/sites/johndunn/2020/12/04/how-cloudflare-became-the-most-important-internet-company-nobody-has-heard-of/?sh=8b7a1c977cfb*

[2] Sousa, J. P. C., Oliveira, L. S., Caldas, D. M. (2021). Vulnerability assessment of fraudulent auction sites - Bug Bounty Justice (*Original title: \Análise de vulnerabilidades em sites fraudulentos de leilões - Bug Bounty Justice*). InterForensics 2021 - Conferência Internacional de Ciências Forenses. Curitiba/PR, Brazil.

[3] Cloudflare (2022) What is a reverse proxy? Proxy Servers explained. *https://www.cloudflare.com/learning/cdn/glossary/reverse-proxy*

[4] Nginx (2022) What is a Reverse Proxy Server? *https://www.nginx.com/resources/glossary/reverse-proxy-server/*

[5] Baloch, Rafay. Ethical hacking and penetration testing guide. CRC Press, 2017.

[6] Cloudflare (2021) Managing DNS records in Cloudflare *https://support.cloudflare.com/hc/en-us/articles/360019093151-Managing-DNS-records-in-Cloudflare*

[7] Record Future (2022) Security Trails: Data Security, Threat Hunting, and Attack Surface Management Solutions for Security Teams *https://securitytrails.com/*

[8] Shodan (2022) Shodan Search Engine *https://www.shodan.io/*

[9] Censys. Censys Search API (2022) *https://search.censys.io/account/api*

[10] Cloudflare (2021) Configuring IP Access Rules *https://support.cloudflare.com/hc/en-us/articles/217074967-Configuring-IP-Access-Rules*

[11] GitHub - m0rtem (2022). CloudFail *https://github.com/m0rtem/CloudFail*

[12] GitHub - christophetd (2022). CloudFlair *https://github.com/christophetd/CloudFlair*

[13] Ip IO. Comprehensive IP address Data, IP geolocation API and database | IPinfo.io (2022) *https://ipinfo.io/*

[14] GitHub - leosol/cloud-feio (2022) Cloud-Feio (Ugly - in portuguese) *https://github.com/leosol/cloud-feio*

[15] Cloudflare (2022) Encryption Modes - Cloudflare SSL docs *https://developers.cloudflare.com/ssl/origin-configuration/ssl-modes*