

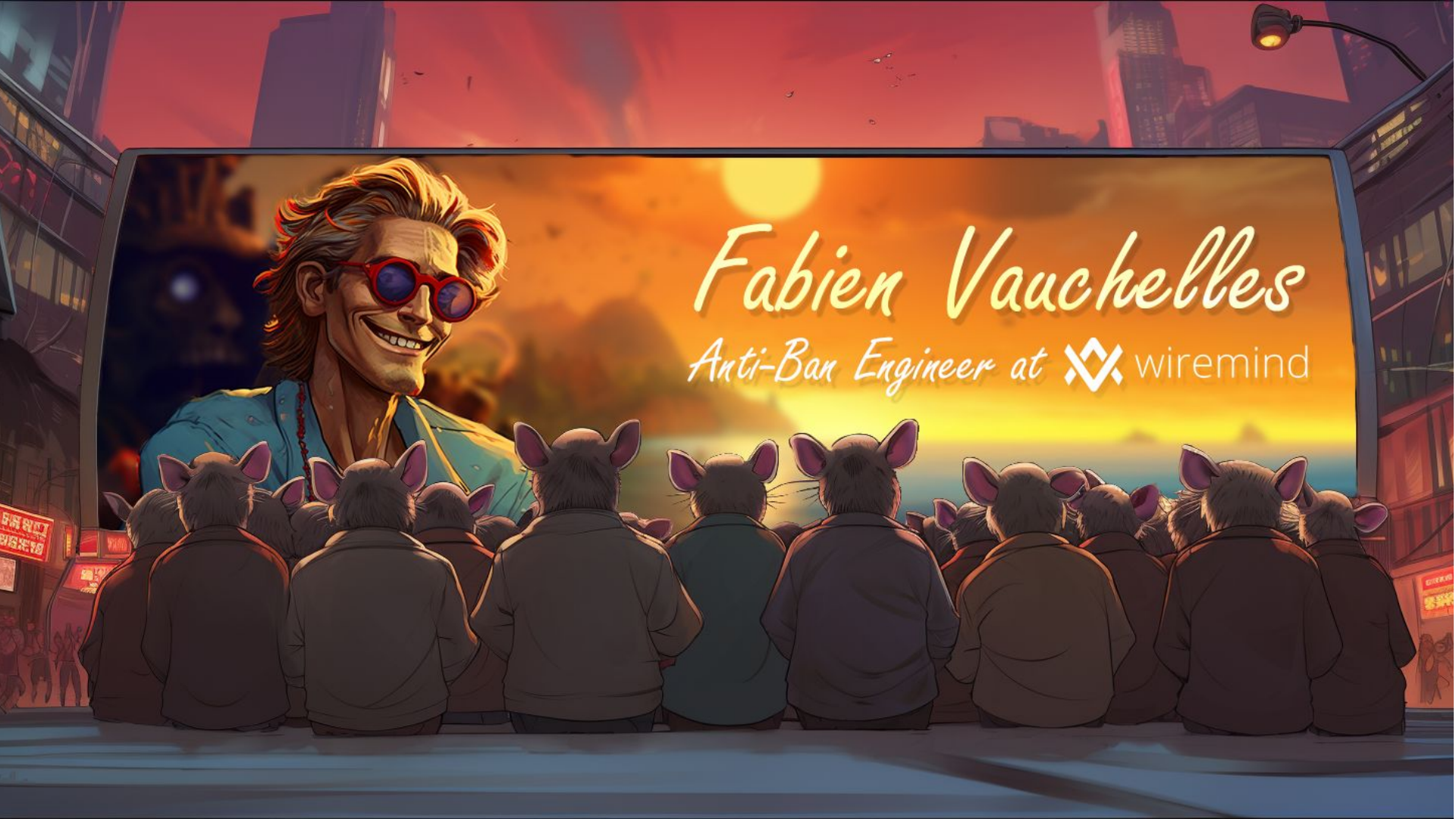
ANATOMY OF

AN ANTIBOT PROTECTION

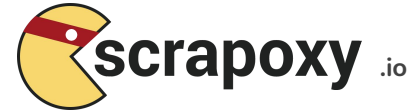


Web Data
Extraction Summit
2023





Fabien Vauchelles
Anti-Ban Engineer at  *wiremind*



- # Open Source
- # Compatible with Cloud Provider
AWS, Azure, GCP, OVH, Digital Ocean, ...
- # Compatible with Proxies Services
Zyte, Rayobyte, IPRoyal, Proxyrack, ...
- # Compatible with Hardware Providers
Proxidize
- # Auto-Rotate proxies to change the IP
address
- # Auto-Scaling proxies to optimize costs
- # On-the-fly request and response
rewriting
- # Stickies sessions with Browser support
- # Distributed architecture

And many cool features to come!

A stylized illustration of a grey mouse with large, pinkish-red ears, wearing red sunglasses and a brown jacket. The mouse is looking towards the right. In the background, a city skyline is visible across a body of water, with a prominent tower resembling the CN Tower. The sky is filled with soft, orange and pink clouds, suggesting a sunset or sunrise. A speech bubble is positioned above the mouse's head, containing the text "I love to play".

I love to play

The Idea



The Idea

I've got an idea.



The Idea

**I've got an idea.
Let's predict job swap.**



The Idea

An illustration of two anthropomorphic mice sitting at a round wooden table on a balcony. The mouse on the left is light grey, wearing a white shirt, a blue tie, and orange-rimmed sunglasses. The mouse on the right is dark grey, wearing a blue suit jacket over a white shirt and orange-rimmed sunglasses. They are both holding white coffee cups. The balcony has a railing, and in the background, there is a cityscape with domed buildings under a blue sky with clouds. A potted plant with orange flowers is on the left.

**I've got an idea.
Let's predict job swap.**

We need data.

The Idea

An illustration of two anthropomorphic mice sitting at a round wooden table on a balcony. The mouse on the left is wearing a light grey shirt and a dark tie, and the mouse on the right is wearing a dark blue jacket over a light shirt. Both are wearing orange-rimmed sunglasses. They are both holding white coffee cups. The background shows a cityscape with domed buildings and a blue sky with clouds. A potted plant with orange fruit is on the left.

**I've got an idea.
Let's predict job swap.**

**We need data.
Million of data.**

The Idea



The Idea



The Idea



The Idea



The Idea



The Idea



I will block you.

Beat the Turing Test



Beat the Turing Test

A woman with curly hair, wearing a green beanie, glasses, and a pink hoodie, is sitting on a couch and typing on a laptop. The room is dimly lit with warm ambient lighting, including a desk lamp and string lights. There are plants and framed pictures in the background.

“Hello, I’m a human living in Paris

Beat the Turing Test

A woman with curly hair, wearing a green beanie, glasses, and a pink hoodie, is sitting on a couch and typing on a laptop. She is in a dimly lit room with warm ambient lighting. In the background, there is a large potted plant, a bookshelf, and some framed art on the wall. A desk lamp is visible behind her, casting a soft glow.

“Hello, I’m a human living in Paris

and I’m excited to be part of the Extract Summit in Dublin.

Beat the Turing Test



“Hello, I’m a human living in Paris

and I’m excited to be part of the Extract Summit in Dublin.

I’m buying a round-trip ticket, during the evening,

Beat the Turing Test



“Hello, I’m a human living in Paris

and I’m excited to be part of the Extract Summit in Dublin.

I’m buying a round-trip ticket, during the evening,

using Safari on my Mac, and I use a DSL connection.”

Beat the Turing Test

“Hello, I’m a human living in Paris [geo]

and I’m excited to be part of the Extract Summit in Dublin.

I’m buying a round-trip ticket, during the evening [time]
[behavior]

using Safari on my Mac, and I use a DSL connection.”
[browser] [os] [ip_type]

Beat the Turing Test

“I’m Scrapy, living in datacenter in Ireland”



Beat the Turing Test



“I’m Scrapy, living in datacenter in Ireland

and I’m not interested in the Extract Summit.

Beat the Turing Test



"I'm Scrapy, living in datacenter in Ireland

and I'm not interested in the Extract Summit.

My goal is to retrieve thousands prices from many airlines.

Beat the Turing Test



“I’m Scrapy, living in datacenter in Ireland

and I’m not interested in the Extract Summit.

My goal is to retrieve thousands prices from many airlines.

I’m accessing to data with a deep link at 2am,

Beat the Turing Test



"I'm Scrapy, living in datacenter in Ireland

and I'm not interested in the Extract Summit.

My goal is to retrieve thousands prices from many airlines.

I'm accessing to data with a deep link at 2am,

using a Python library within my Docker.

Beat the Turing Test



"I'm Scrapy, living in datacenter in Ireland

and I'm not interested in the Extract Summit.

My goal is to retrieve thousands prices from many airlines.

I'm accessing to data with a deep link at 2am,

using a Python library within my Docker.

My internet provider is AWS

Beat the Turing Test



“I’m Scrapy, living in datacenter in Ireland

and I’m not interested in the Extract Summit.

My goal is to retrieve thousands prices from many airlines.

I’m accessing to data with a deep link at 2am,

using a Python library within my Docker.

My internet provider is AWS

and I love to randomize my User-Agent.”

Beat the Turing Test



“I’m Scrapy, living in datacenter in Ireland [geo]

and I’m not interested in the Extract Summit.

My goal is to retrieve thousands prices from many airlines. [behavior]

I’m accessing to data with a deep link at 2am [time]

using a Python library within my Docker. [browser] [os]

My internet provider is AWS [ip_type]

and I love to randomize my User-Agent.”

Beat the Turing Test



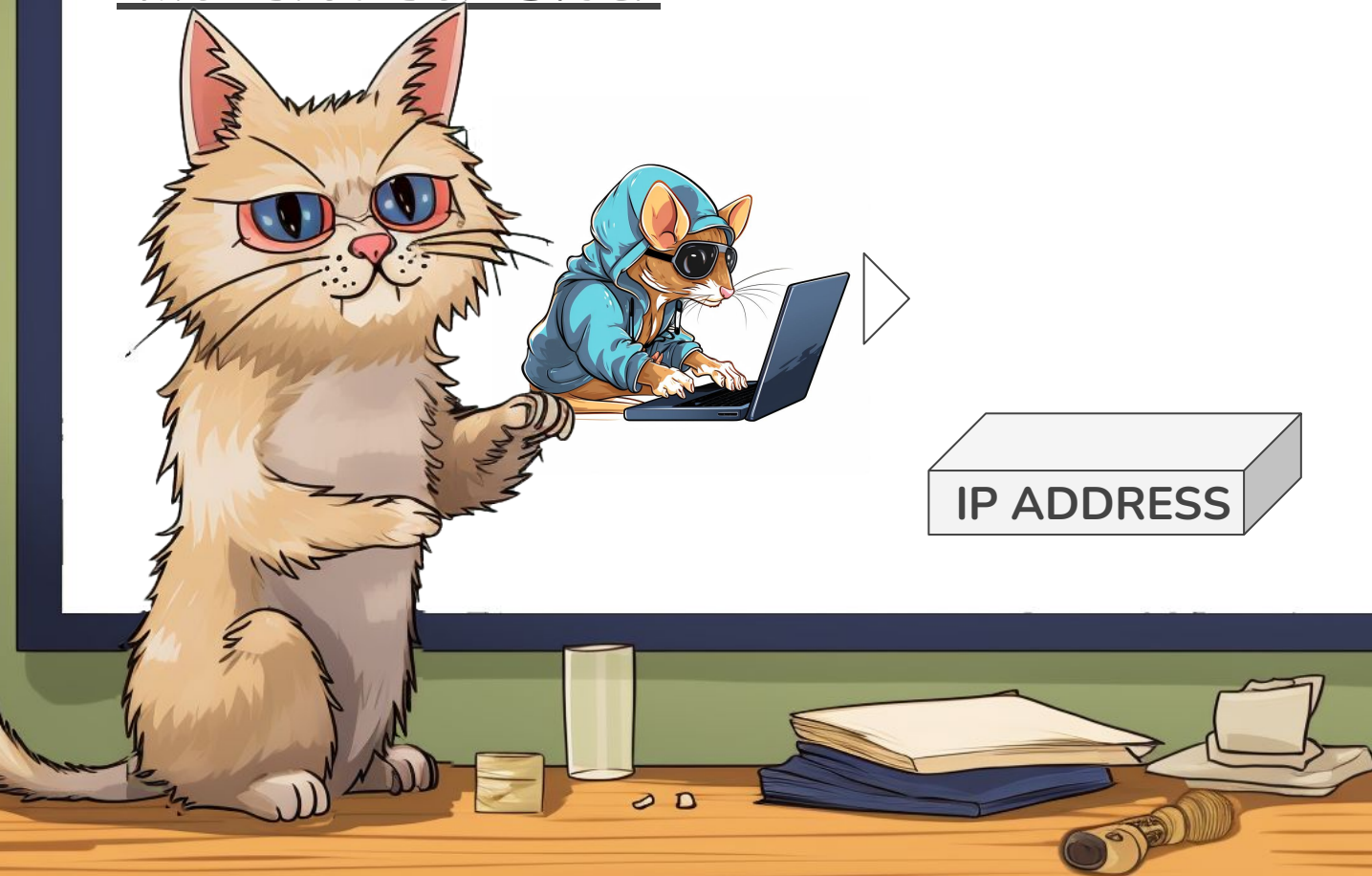
The Browser Stack



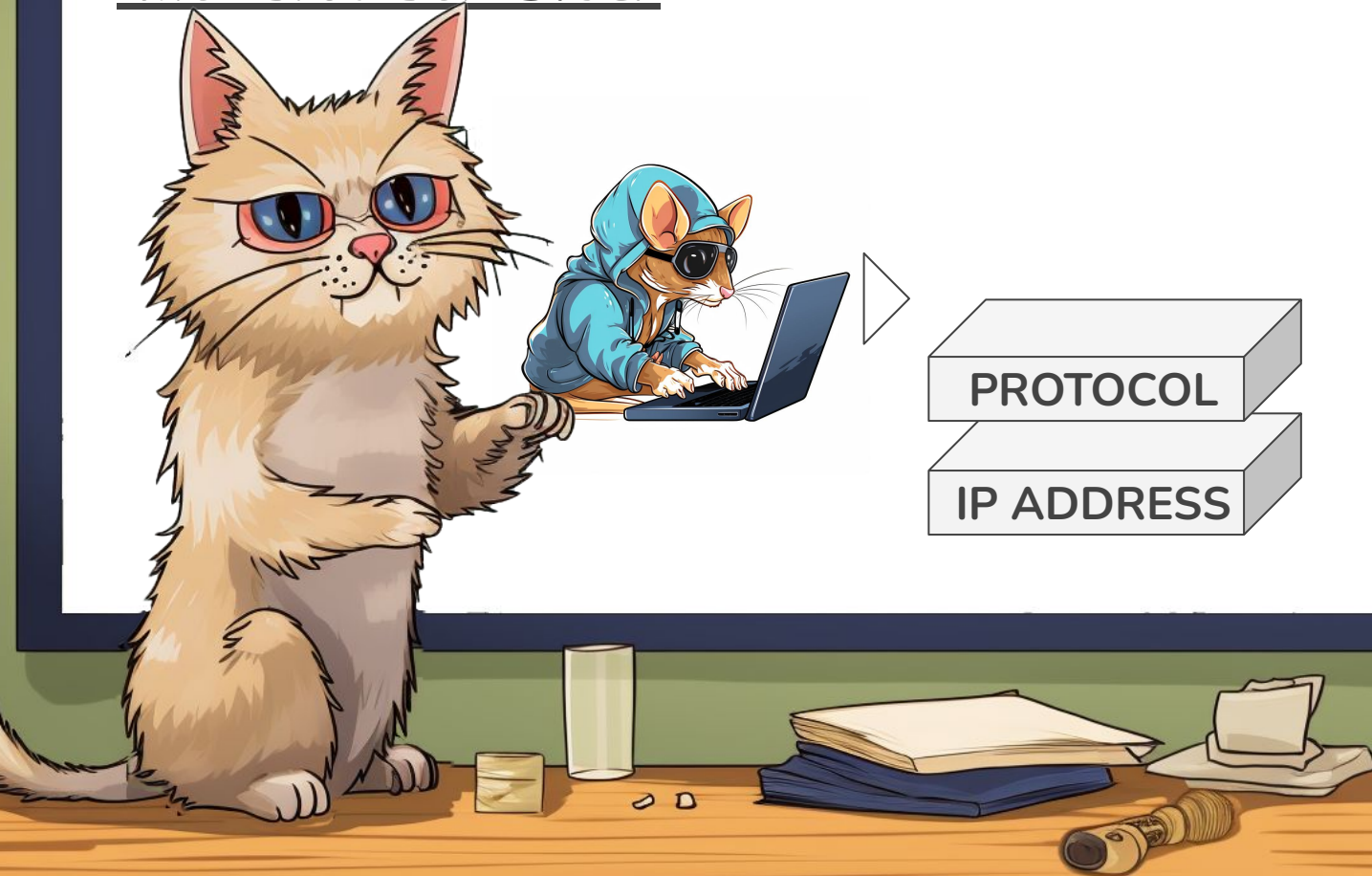
The Browser Stack



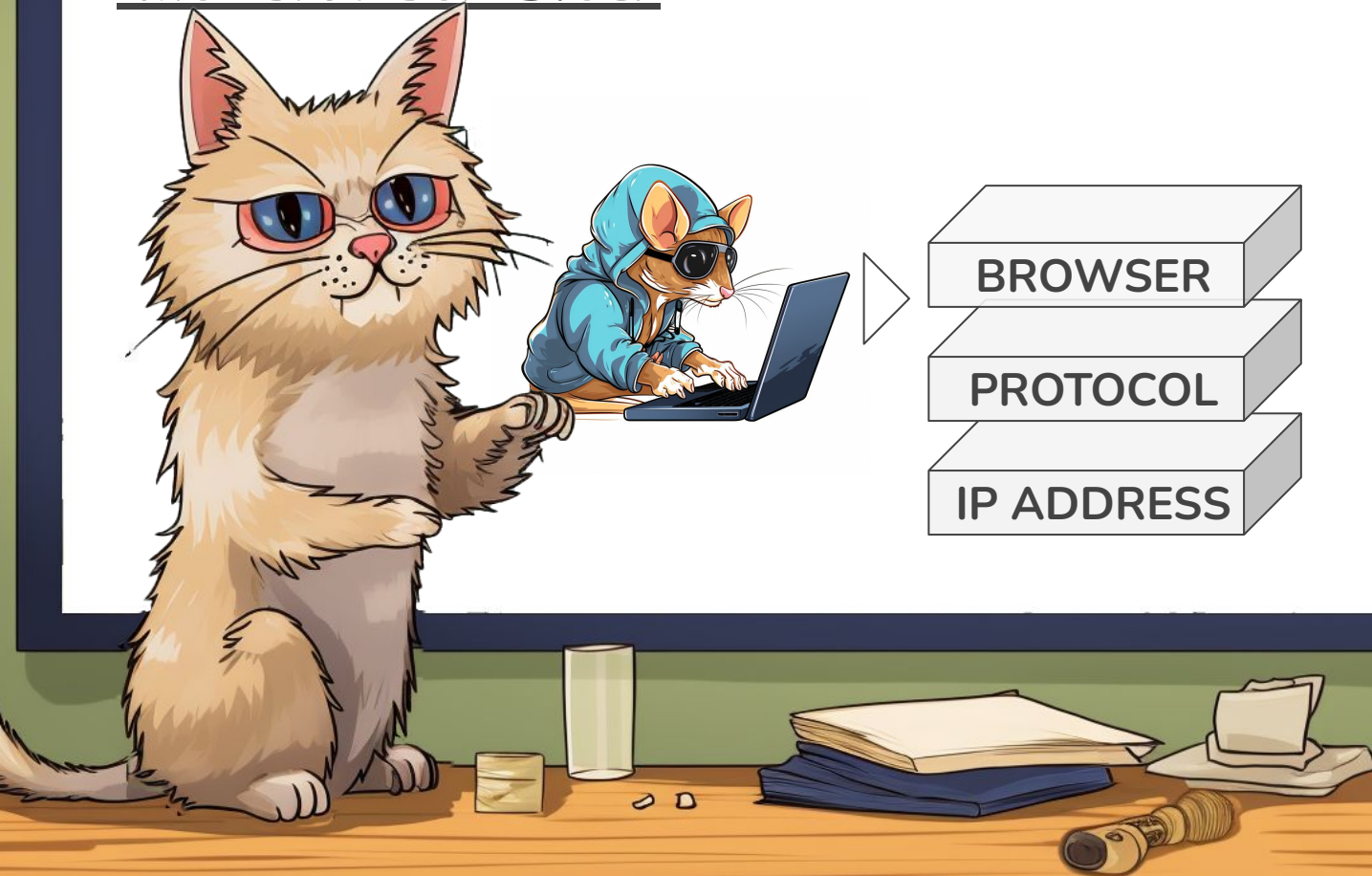
The Browser Stack



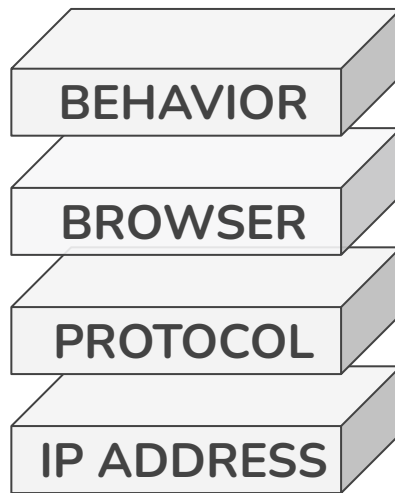
The Browser Stack



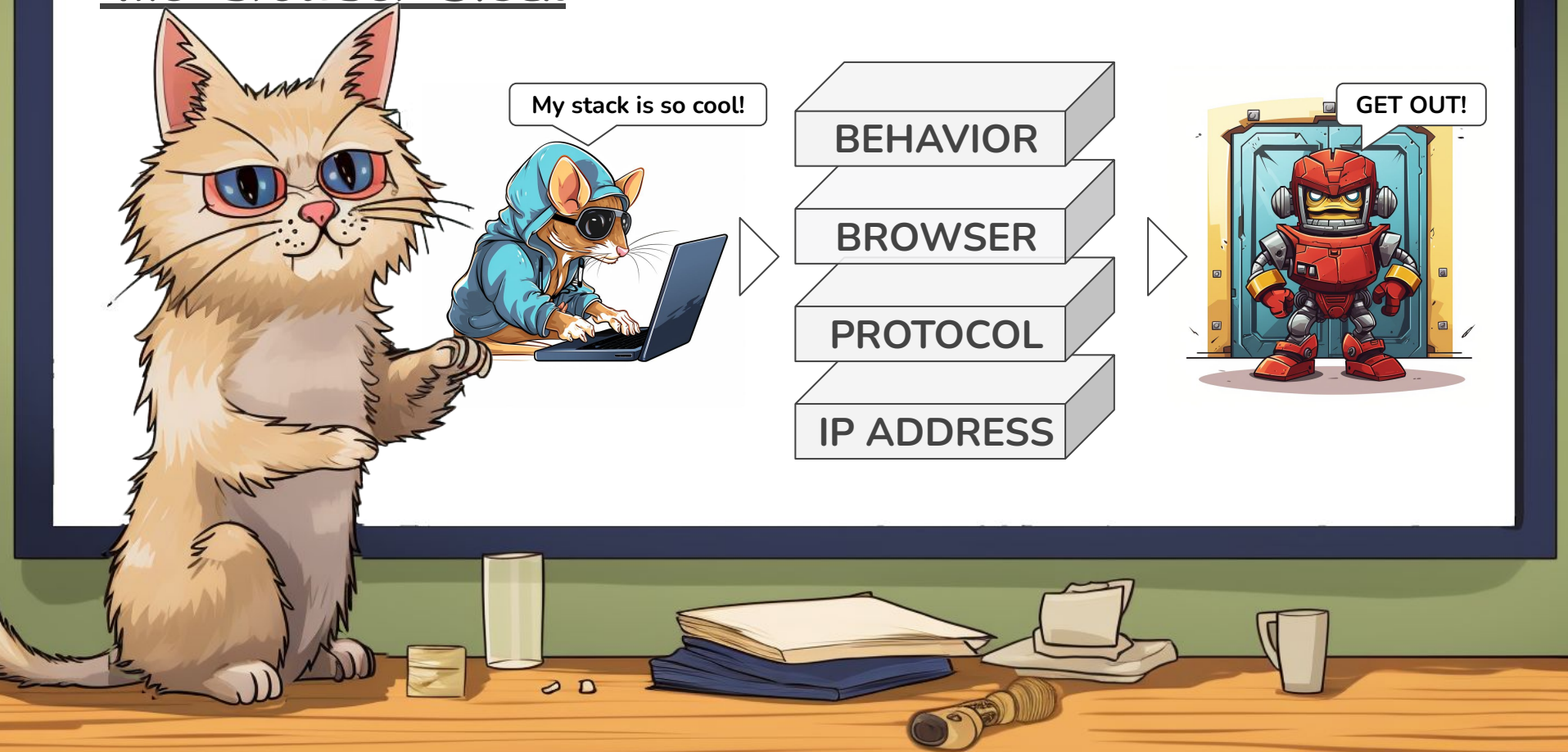
The Browser Stack



The Browser Stack



The Browser Stack



IP Address Layer

ipgeolocation.io

thyme.apnic.net

ipregistry.co

ipinfo.io

ip2location.com

maxmind.com

ipstack.com

github.com/nitefood/asn



Network Protocol Layer



TCP/IP headers

TLS protocol

HTTP/2 headers



TCP/IP Headers



	Windows	Linux 2.2	Linux 2.0
Window size			
Initial TTL			
Max segment size			
DF flag (Don't Fragment)			
Window scaling value			
sackOK flag			
nop flag			
Initial Packet Size			

from "Security warrior", o'Reilly, Cyrus Peikari & Anton Chuvaki



TCP/IP Headers



	Windows	Linux 2.2	Linux 2.0
Window size	8192	32120	512
Initial TTL	128	64	64
Max segment size	1460	1460	1460
DF flag (Don't Fragment)	1	1	0
Window scaling value	0	0	0
sackOK flag	0	1	0
nop flag	0	1	0
Initial Packet Size	44	60	44

from "Security warrior", o'Reilly, Cyrus Peikari & Anton Chuvaki



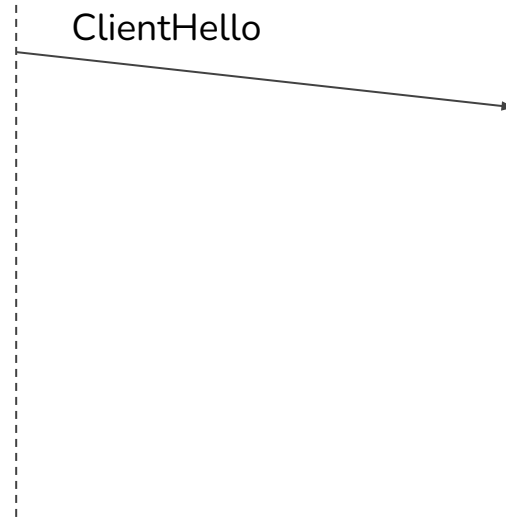
TLS / Handshake Protocol



TLS / Handshake Protocol



ClientHello



TLS / Handshake Protocol



ClientHello

ServerHello

Key Exchange



TLS / Handshake Protocol



ClientHello

TLS version	Cipher	list of Extensions
769,	47,	65281-0-11-35-5-16



ClientHello

ServerHello

Key Exchange



TLS / Handshake Protocol



ClientHello

TLS version	Cipher	list of Extensions
769,	47,	65281-0-11-35-5-16

=

JA3 Fingerprint



ClientHello

ServerHello

Key Exchange

HTTP/2 Headers

Browserleaks.com

SETTINGS Frame :

Length	24
Settings	SETTINGS_HEADER_TABLE_SIZE: 65536
	SETTINGS_ENABLE_PUSH: 0
	SETTINGS_INITIAL_WINDOW_SIZE: 6291456
	SETTINGS_MAX_HEADER_LIST_SIZE: 262144

WINDOW_UPDATE Frame :

Length	4
Window Size Increment	15663105

HEADERS Frame :

Length	357
Stream ID	1
Headers	:method
	:authority
	:scheme
	:path
	sec-ch-ua
	sec-ch-ua-mobile
	user-agent
	sec-ch-ua-platform
	accept
	origin



HTTP Layer



HTTP Layer

I'm a chrome
browser...



HTTP Layer



HTTP Headers



```
function getHeadersListFingerprint(request) {  
    return [...Object.keys(request.headers)].join(',');  
}
```



Browser Layer



JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS



Browser Layer



device
fo.me

Date & Time:

System (Live):

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 GMT+2) (DST: Yes)

System Time Zone:

Europe/Paris

Local (Live):

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 CEST) (DST: Yes)

Local Time Zone:

Europe/Paris

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS



Browser Layer



device
Fo.me

Date & Time:

System (Live):

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 GMT+2) (DST: Yes)

System Time Zone:

Europe/Paris

Local (Live):

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 CEST) (DST: Yes)

Local Time Zone:

Europe/Paris

Graphics Card Name / Driver:

NVIDIA, NVIDIA GeForce RTX 3050 Ti Laptop GPU
Direct3D11 vs_5_0 ps_5_0, D3D11

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

JS

Browser Layer

JS

JS

Date & Time:

System (Live):

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 GMT+02:00 (DST: Yes))

System Time Zone:

Europe/Paris

Local (Live):

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 CEST) (DST: Yes)

Local Time Zone:

Europe/Paris

Canvas:

Supported

Canvas Fingerprinting:

Allowed

JS

Graphics Card Name / Driver:

NVIDIA, NVIDIA GeForce RTX 3050 Ti Laptop GPU
Direct3D11 vs_5_0 ps_5_0, D3D11

JS

device
Fo.me



Browser Layer

JS

JS

Number of Speakers:

6

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 GMT+02:00 (DST: Yes))

System Time Zone:

Europe/Paris

Local (Live):

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 CEST) (DST: Yes)

Local Time Zone:

Europe/Paris

Canvas:

Supported

Canvas Fingerprinting:

Allowed

JS

Graphics Card Name / Driver:

NVIDIA, NVIDIA GeForce RTX 3050 Ti Laptop GPU
Direct3D11 vs_5_0 ps_5_0, D3D11

JS

device
Fo.me



Browser Layer

JS

JS

Number of Speakers:

6

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00 GMT+02:00 (DST))
Yes)

Canvas:

Speaker 1:

Par défaut - Speakers (Dell USB Audio) (17e9:6006)

gerprinting:

Speaker 2:

Communications - Headset Earphone (HyperX Virtual Surround Sound) (0951:16a4)

JS

Speaker 3:

Speakers (Dell USB Audio) (17e9:6006)

ard Name / Driver:

NVIDIA GeForce RTX 3050 Ti Laptop GPU
Direct3D11 vs_5_0 ps_5_0, D3D11

JS

device
Fo.me



Browser Layer



device JS
Fo.me

Number of Speakers:

6

Thu, Oct 12, 2023, 17:17:19 (UTC+02:00)
Yes

Speaker 1:

Par défaut - Speakers (Dell USB Audio) (17e9

Speaker 2:

Communications - Headset Earphone (HyperX
Surround Sound) (0951:16a4)

Speaker 3:

Speakers (Dell USB Audio) (17e9:6006)

Microphones:

Microphone Permission:

Granted

Number of microphones:

5

Microphone Labels:

Microphone 1:

Par défaut - Microphone (2- RODE NT-USB) (19f7:0003)

Card Name / Driver:

NVIDIA GeForce RTX 3050 Ti Laptop GPU
Direct3D11 vs_5_0 ps_5_0, D3D11

Browser Layer



Device
Fo.me

Number of Speakers:

6

Cameras:

Camera Permission:

Not granted, or denied. Permission required to detect
Par dé correct number of devices and devices' labels.

Speakers:

Number of cameras:

1

Speakers:

Camera Labels:

Speakers:

Camera 1:

Camera 1

Microphones:

Microphone Permission:

Microphones:

Labels:

Microphone (2- RODE NT-USB) (19f7:0003)

Name / Driver:

GeForce RTX 3050 Ti Laptop GPU

5_0 ps_5_0, D3D11

Browser Layer



Device Fo.me

Number of Speakers:

6

Cameras:

Camera Permission:

Not granted, or denied. Per
Par dé correct number of devices :

Speakers:

Comm

Surrou

Speakers:

Speake

Number of cameras:

1

Camera Labels:

Camera 1:

Camera 1

Microphones:

Microphone Permission:

Microphones:

Battery Status (Live):

Level:

100%

Charging:

Yes

Time remaining before charged:

0 sec

ne (2- RODE NT-USB) (19f7:0003)

/ Driver:

orce RTX 3050 Ti Laptop GPU

_5_0, D3D11



Browser Layer



device
Fo.me

Number of Speakers:

6

Cameras:

Ye

Camera Permission:

Speak Not granted, or denied. Per
Par dé correct number of devices :

Number of cameras:

Speak
Comm
Surrou

1

Camera Labels:

Speak

Camera 1:

Camera 1

Microphones:

Microphone Permission:

Fonts List: [Show / Hide ↓](#)

Agency FB, Algerian, Arial, Arial Black, Arial Narrow, Arial Rounded MT, Bahnschrift, Baskerville Old Face, Bauhaus 93, Bell MT, Berlin Sans FB, Bernard MT Condensed, Blackadder ITC, Bodoni 72, Bodoni 72 Oldstyle, Bodoni 72 Smallcaps, Bodoni MT, Bodoni MT Black, Bodoni MT Condensed, Book Antiqua, Bookman Old Style, Bookshelf Symbol 7, Bradley Hand ITC, Britannic, Broadway, Brush Script MT, Calibri, Calibri Light, Californian FB, Calisto MT, Cambria, Cambria Math, Candara, Candara Light, Castellar, Caveat, Centaur, Century, Century Gothic, Century Schoolbook, Chiller, Colonna MT, Comic Sans MS, Consolas, Constantia, Cooper, Cooper Black, Copperplate Gothic,

Battery Status

Level:

100%

Charging:

Yes

Time remaining

0 sec

Browser Layer



device
Fo.me

Number of Speakers:

6

Cameras:

TI
Ye

Browser Plugins:

Hide / Show ↓

Speakers:

Not granted

Par défaut correcteur

Speakers:

Comm

Surround

Speakers:

Speakers

Name: PDF Viewer

Version: Unknown. Could not detect.

Description: Portable Document Format

Filename: internal-pdf-viewer

Camera List:

Name: Chrome PDF Viewer

Version: Unknown. Could not detect.

Description: Portable Document Format

Filename: internal-pdf-viewer

Microphones:

Microphone Permission:

Fonts List: Show / Hide ↓

Agency FB, Algerian, Arial, Arial Black, Arial Narrow, T, Bahnschrift, Baskerville Old Face, MT, Berlin Sans FB, Bernard MT, kadder ITC, Bodoni 72, Bodoni 72, 72 Smallcaps, Bodoni MT, Bodoni MT, Condensed, Book Antiqua, Bookman, elf Symbol 7, Bradley Hand ITC, ay, Brush Script MT, Calibri, Calibri, FB, Calisto MT, Cambria, Cambria, Sandara Light, Castellar, Caveat, Century Gothic, Century Schoolbook, MT, Comic Sans MS, Consolas, er, Cooper Black, Copperplate Gothic,

Browser Layer



device
Fo.me

Number of Speakers:

6

Cameras:

TI
Ye

Browser Plugins:

Hide / Show ↓

Speakers:

Not granted

Par défaut correcteur

Speakers:

Comm

Surround

Speakers:

Surround

Surround

Camera List:

Camera 1

Camera 1

Camera 1

Camera 1

Camera 1

Name: PDF Viewer

Version: Unknown. Could not detect.

Description: Portable Document Format

Filename: internal-pdf-v

Name: Chrome PDF Vie

Version: Unknown. Could not detect.

Description: Portable Document Format

Filename: internal-pdf-viewer

Microphones:

Microphone Permission:

Fonts List:

Show / Hide ↓

Agency FB, Algerian, Arial, Arial Black, Arial Narrow, T, Bahnschrift, Baskerville Old Face, MT, Berlin Sans FB, Bernard MT, kadder ITC, Bodoni 72, Bodoni 72, 72 Smallcaps, Bodoni MT, Bodoni MT, Condensed, Book Antiqua, Bookman, half Symbol 7, Bradley Hand ITC, ush Script MT, Calibri, Calibri, alisto MT, Cambria, Cambria, a Light, Castellar, Caveat, ury Gothic, Century Schoolbook, MT, Comic Sans MS, Consolas, er, Cooper Black, Copperplate Gothic,

Ad Blocker:

None detected

Browser Layer



device
Fo.me

Number of Speakers:

6

WebRTC:

Enabled

Microphones:

Microphone Permission:

Fonts List:

Show / Hide ↓

Agency FB, Algerian, Arial, Arial Black, Arial Narrow, T, Bahnschrift, Baskerville Old Face, MT, Berlin Sans FB, Bernard MT, kadder ITC, Bodoni 72, Bodoni 72, 72 Smallcaps, Bodoni MT, Bodoni MT, Condensed, Book Antiqua, Bookman, elf Symbol 7, Bradley Hand ITC, ush Script MT, Calibri, Calibri, alisto MT, Cambria, Cambria, a Light, Castellar, Caveat, ury Gothic, Century Schoolbook, MT, Comic Sans MS, Consolas, er, Cooper Black, Copperplate Gothic,

Cameras:

Camera P

Not granted

Par défaut correcteur

Speakers:

Comm

Surrou

Speakers:

Speakers

Browser Plugins:

Hide / Show ↓

Name: PDF Viewer

Version: Unknown. Could not detect.

Description: Portable Document Format

Filename: internal-pdf-v

Ad Blocker:

None detected

Name: Chrome PDF Vie

Version: Unknown. Could not detect.

Description: Portable Document Format

Filename: internal-pdf-viewer

Browser Layer



device
Fo.me

Number of Speakers:

6

WebRTC:

Enabled

Microphones:

Micronphone Permission:

Fonts List:

Show / Hide ↓

Agency FB, Algerian, Arial, Arial Black, Arial Narrow, T, Bahnschrift, Baskerville Old Face, MT, Berlin Sans FB, Bernard MT, kadder ITC, Bodoni 72, Bodoni 72, 72 Smallcaps, Bodoni MT, Bodoni MT, Condensed, Book Antiqua, Bookman, elf Symbol 7, Bradley Hand ITC, ush Script MT, Calibri, Calibri, alisto MT, Cambria, Cambria, a Light, Castellar, Caveat, ury Gothic, Century Schoolbook, MT, Comic Sans MS, Consolas, er, Cooper Black, Copperplate Gothic,

Cameras:

Camera P:

Not grante

Par dé correct nur

Speak

Par dé

Speak

Comm

Surrou

Speak

Speake

Browser Plugins:

Hide / Show ↓

Name

Versic

Descr

Filena

Name

Camera 1:

Camera 1

Browser Window Size:

Outer:

3072 x 1680 (pixels)

Inner:

1763 x 1571 (pixels)


Description: Portable Document Format

Filename: internal-pdf-viewer

Browser Layer



Device Info HW	
GENERAL	SOC
SYSTEM	
SCREEN	
MEMORY	
CPU	
Model	ranchu
Cores	4
Family	Family 6 Model 6 Stepping 3
Machine	x86_64
ABI	x86_64
Clock speed	1 - 2 MHz
Governor	schedutil
Cache L3	16 MiB
Supported ABI	x86_64 arm64-v8a
GPU	
Android Emulator OpenGL ES Translator (Mesa Intel(R) UHD Graphics (TGL GT1))	
Vendor	Google (Intel)
OpenGL ES	3.1 (4.6 (Core Profile) Mesa 23.0.4-0ubuntu1~23.04.1)
Vulkan	1.1
Extensions	52

Device Info HW	
GENERAL	SOC
SYSTEM	
SCREEN	
MEMORY	
 sdk_gphone_x86_64	
Manufacturer	Google
Brand	google
Model	sdk_gphone_x86_64
Release	13
API	33
Codename	Tiramisu
Device	emu64xa
Product	sdk_gphone_x86_64
Board	goldfish_x86_64
Build	TE1A.220922.031
Java VM	ART 2.1.0
Security	05.11.2022
Baseband	1.0.0.0
GPS	Android Studio Emulator GPS / 2020
Bluetooth	4.x
Build type	userdebug
Tags	dev-keys



Behavior Layer



Behavior Layer

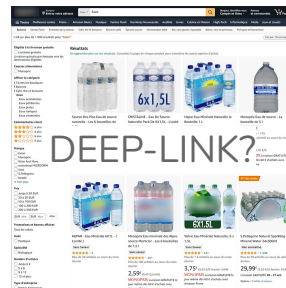


SEARCH QUERY?



OR

DEEP-LINK?



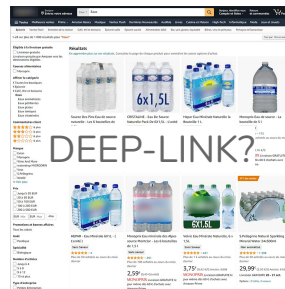
Behavior Layer



SEARCH QUERY?



OR



DEEP-LINK?

Full Content?

HTML



CSS



OR

Only HTML?

HTML

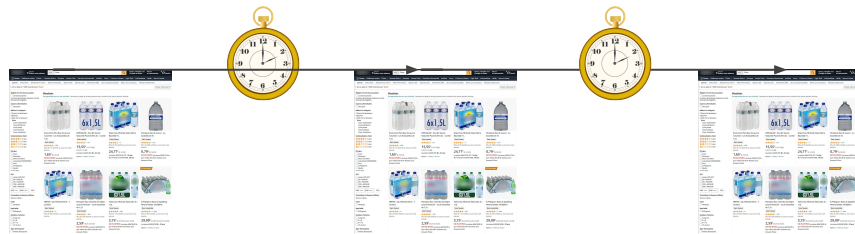


CSS



Behavior Layer

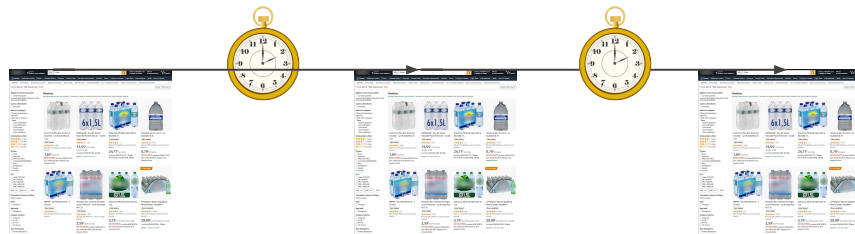
Interval between webpages



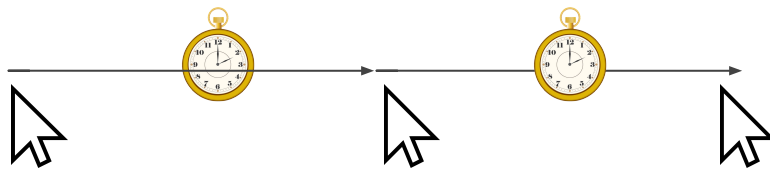
Behavior Layer



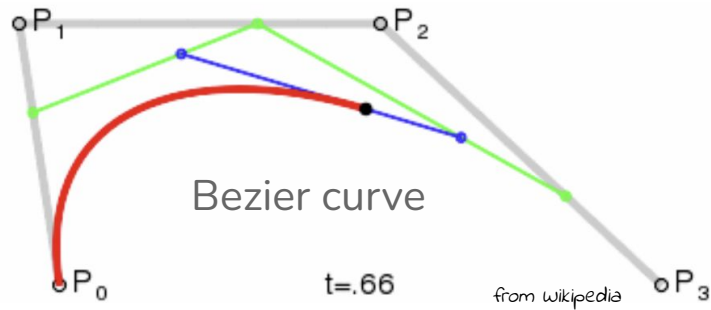
Interval between webpages



Delay between mouse clicks



Behavior Layer





CAPTCHA!

AI Session Reputation



Fracture



**TOO MUCH
CHALLENGES!**

False Positives are the key



False Positives are the key



False Positives are the key



False Positives are the key



Collecting millions of clean records is too challenging and too expensive...

False Positives are the key



They focus on the most effective signals.

False Positives are the key



IP Address

HTTP headers


Javascript

Behavior





Analysis Method / Browser






 **Wappalyzer** 🔍 ⚙️ 🔄



[TECHNOLOGIES](#) [PLUS D'INFORMATION](#) [Export](#)


Outil de statistiques
 [Branch](#)


Framework JavaScript
 [React](#)


Outil de suivi de problèmes
 [GetFeedback](#)


Sécurité
 [DataDome](#)
 [HSTS](#)

Librairies JavaScript
 [Swipe](#)
 [Loadable-Components](#)


PaaS
 [Amazon Web Services](#)



Frameworks UI
 [Bootstrap](#) 5



Conformité des cookies
 [OneTrust](#)


 **Wappalyzer** 🔍 ⚙️ 🔄


[TECHNOLOGIES](#) [PLUS D'INFORMATION](#) [Export](#)


Outil de statistiques
 [Dynatrace](#)


Sécurité
 [Akamai Bot Manager](#)
 [HSTS](#)

Divers
 [PWA](#)
 [Open Graph](#)

Frameworks UI
 [Bootstrap](#) 3.4.1

A/B testing
 [Monetate](#)

Personnalisation
 [Monetate](#)

RUM
 [Boomerang](#)



Analysis Method



Elements Console Sources **Network** Performance Memory Application Security Lighthouse >> 21

⏏ ⛔ 🔍 ☒ Preserve log ☐ Disable cache No throttling 📶 ⬆️ ⬇️

Filter ☐ Invert ☐ Hide data URLs ☐ Hide extension URLs All Fetch/XHR JS CSS Img Media Font Doc WS Wasm

10000 ms 20000 ms 30000 ms 40000 ms 50000 ms 60000 ms 70000 ms 80000 ms 90000 ms 100000 ms 110000 ms 12

Name

- ↳ ajax.php
- ↳ icon
- ↳ dependencies/default.css
- ↳ webpackruntime/hot/hotModuleReplacement.js
- ↳ webfontloading.js, font-loading, & font-loading.async.js
- ↳ phone-number
- ↳ country
- ↳ companies
- ↳ security/permissions/default.js
- ↳ offer
- ↳ ...

Headers Preview Response Initiator Timing Cookies

▼ General

Request URL: <https://www.example.com/dependencies/default.css>

▼ Response Headers ☐ Raw

Access-Control-Allow-Credentials: true

Access-Control-Allow-Origin: <https://www.example.com/>

Connection: keep-alive

Content-Encoding: gzip

Content-Length: 264

Content-Type: application/json;charset=UTF-8

Set-Cookie: ak=...

Analysis Method / Android



Charles
WEB DEBUGGING PROXY



**HTTP
TOOLKIT**

FRIIDA



Anti-bot Architecture



Browser



Website



Anti-bot



Anti-bot Architecture

1. *Gather information*

Browser



Website



Anti-bot



Anti-bot Architecture

1. Gather information

Browser



Website

2. Ask a token



Anti-bot



Anti-bot Architecture

1. Gather information

Browser



2. Ask a token

3. Browse with a token



Website



Anti-bot



Anti-bot Architecture

1. Gather information

Browser



2. Ask a token

3. Browse with a token



Website

4. Ask token's reputation score



Anti-bot



Websites evolve fast



Anti-ban Methodology



Step 1

Code a scraper

Requests Delay
Concurrency
User Agent
Referer

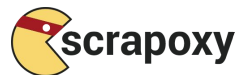
Cookies
TLS version
Payload



Anti-ban Methodology

Step 2 Use Standard Proxies

Datacenter Proxies



Anti-ban Methodology



Step 3

Start a Headless Browser

Puppeteer
Playwright
+ Stealth plugin
+ Ghost cursor



Anti-ban Methodology



Step 4

Use Advanced Proxies

Residential Proxies

IP addresses from home internet users which agreed to share their connection.

Mobile Proxies

IP addresses from 4G/5G USB keys. You don't host the hardware.

ISP Proxies

Datacenter proxies which use IP addresses of personal Internet providers

Hardware Proxies

IP addresses from 4G/5G USB keys. You host the hardware.



Anti-ban Methodology



Step 5

Shift to Headful Browser

Commercial or Open Source

XFVB display

PyAutoGui



Anti-ban Methodology

Step 6 Call an Unblocker API

Zyte API



Anti-ban Methodology



Step 7

Solve Captcha with AI

CapSolver

CapMonster



Anti-ban Methodology

Step 8 Generate Antibot Token

Discord

Telegram



Reverse Engineering



POSTMAN



Magisk



Android Studio

Android Emulator



FRIDA

BABEL



AuroraOSS



mitmproxy

p0f

CreepJS

zadaxt.py

curl-impersonate



OSfooler-NG

Open Source Tools



scrapoxy

Rama

Distortion proxy software
to be anonymous.



Scrapy



crawlee

estela



PyAutoGUI

ghost-cursor



Xvfb



Playwright

Fingerprint

Proxies

Scraping

Automation