Le Browser Fingerprinting,

Une technique de traçage peu connue

Thomas Lamiaux

26/10/2023, Enjeux numériques du monde contemporain

Table of contents

Plan

1. Le principe du browser fingerprinting

1. Le principe du browser fingerprinting

1.1 L'idée derrière le browser fingerprinting

L'envoi automatique de données

L'envoi automatique de données

Votre navigateur (safari, firefox, chrome) envoie automatiquement des informations à toutes les pages web

L'envoi automatique de données

L'envoi automatique de données

Votre navigateur (safari, firefox, chrome) envoie automatiquement des informations à toutes les pages web

Exemples

Données envoyées	Permet
La taille de l'écran	d'afficher la page correctement
La language	dans la bonne langue
Les polices disponibles	dans une police qui existe

L'origine du fingerprinting

Une forme d'empreinte digitale ?

En 2010, Peter Eckersley s'est rendu compte qu'en récupérant seulement 8 de ces données:

L'origine du fingerprinting

Une forme d'empreinte digitale ?

En 2010, Peter Eckersley s'est rendu compte qu'en récupérant seulement 8 de ces données:

- Il est possible de distinguer de manière unique 94% des utilisateurs parmi 500 000 personnes

L'origine du fingerprinting

Une forme d'empreinte digitale ?

En 2010, Peter Eckersley s'est rendu compte qu'en récupérant seulement 8 de ces données:

- Il est possible de distinguer de manière unique 94% des utilisateurs parmi 500 000 personnes
- La plupart de ces données évoluent lentement

L'origine du fingerprinting

Une forme d'empreinte digitale ?

En 2010, Peter Eckersley s'est rendu compte qu'en récupérant seulement 8 de ces données:

- Il est possible de distinguer de manière unique 94% des utilisateurs parmi 500 000 personnes
- La plupart de ces données évoluent lentement
- Il est ainsi possible de les utiliser pour identifier les internautes, sans laisser aucune trace sur l'appareil

L'origine du fingerprinting

Une forme d'empreinte digitale ?

En 2010, Peter Eckersley s'est rendu compte qu'en récupérant seulement 8 de ces données:

- Il est possible de distinguer de manière unique 94% des utilisateurs parmi 500 000 personnes
- La plupart de ces données évoluent lentement
- Il est ainsi possible de les utiliser pour identifier les internautes, sans laisser aucune trace sur l'appareil

Une méthode de pistage

Une empreinte unique peut permettre de pister l'utilisateur entre différent sites

Exemples : langues et fuseaux horaires

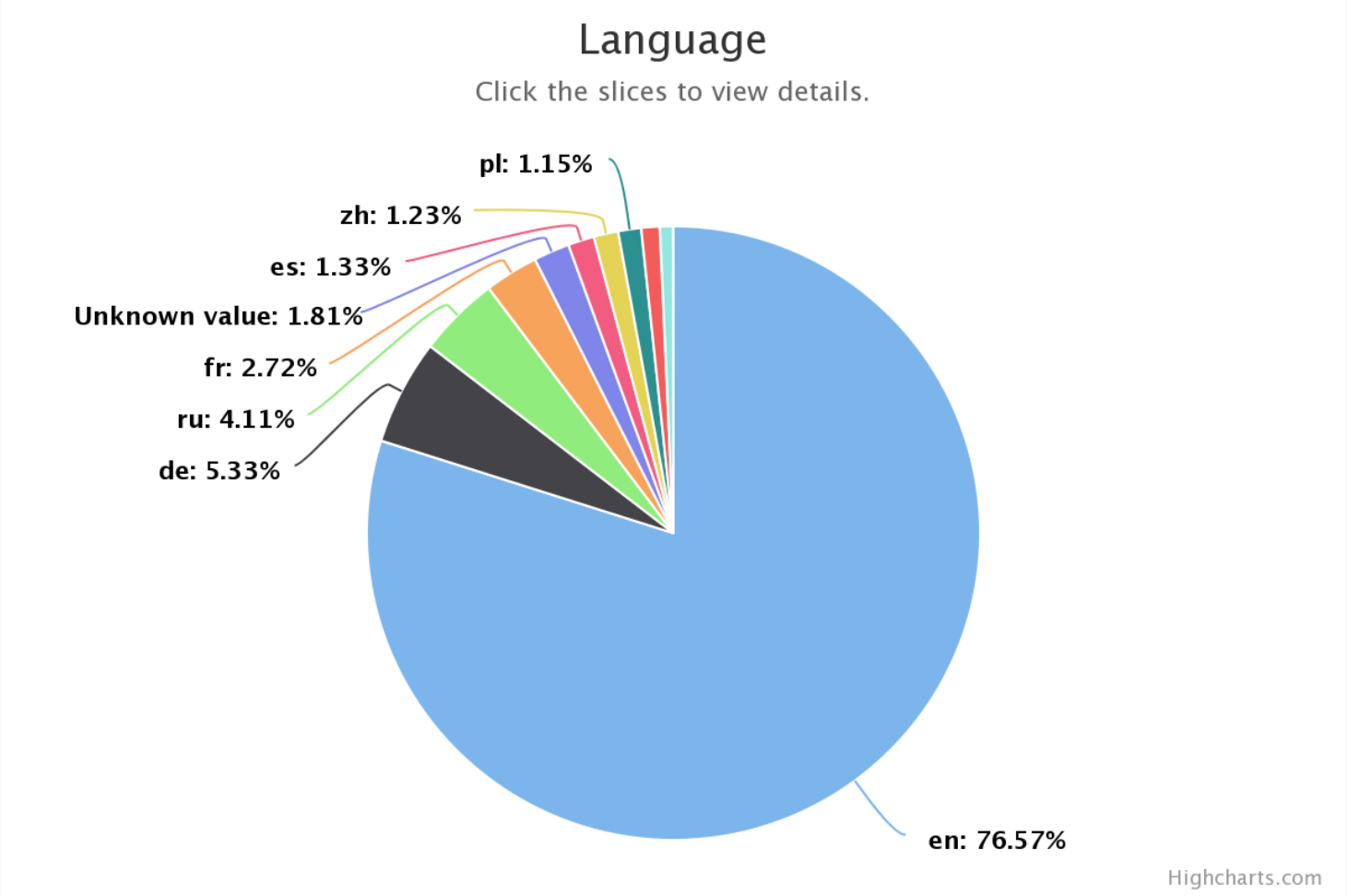

Figure 1: Répartition des langues sur 2.106.504 users,
Le 09/10, de amiunique.org

Exemples : languages et fuseaux horaires

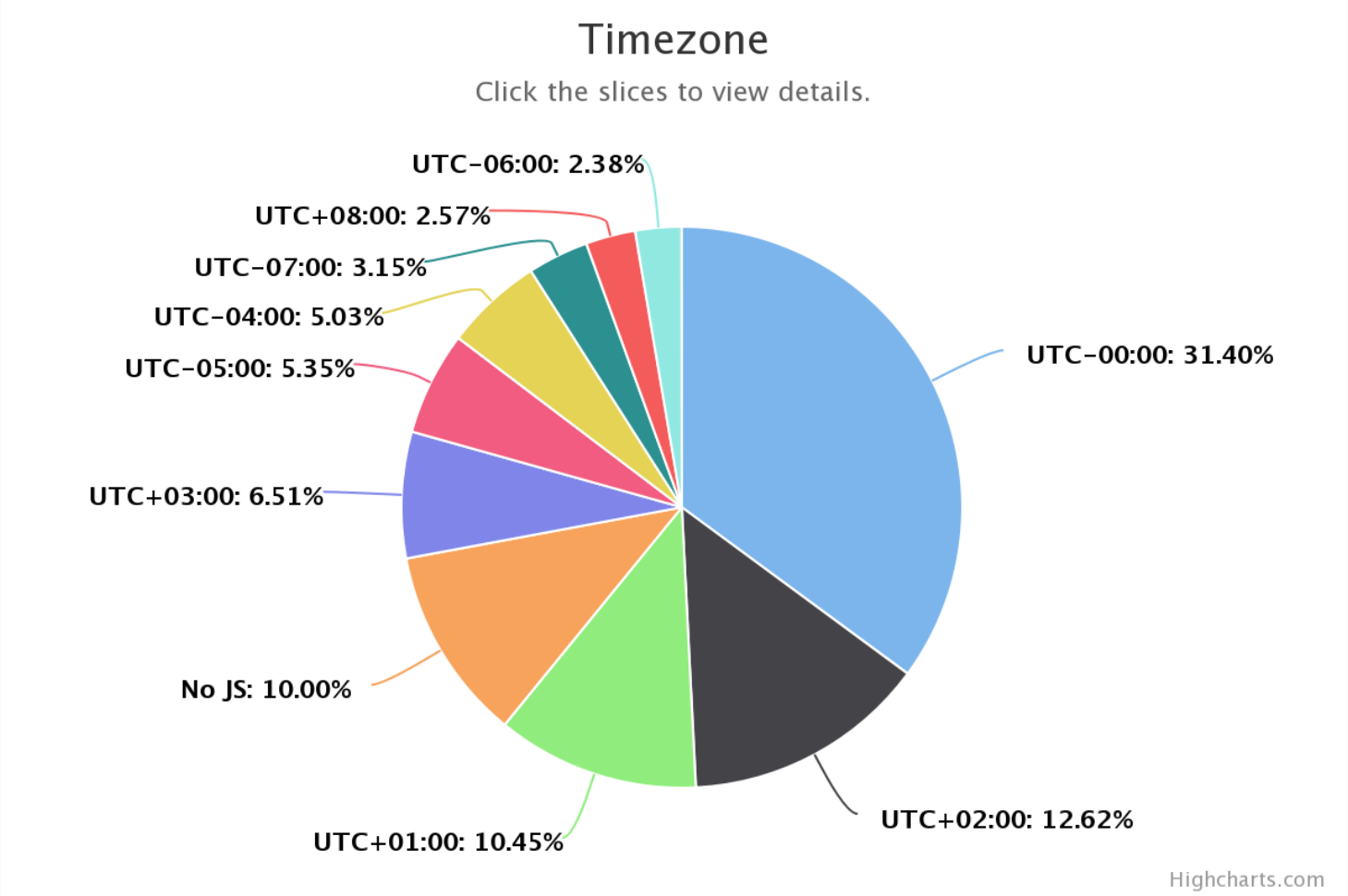

Figure 2: Répartition des fuseaux horaires sur 2.106.504 users, Le 09/10, de amiunique.org

Exemples : OS et navigateurs

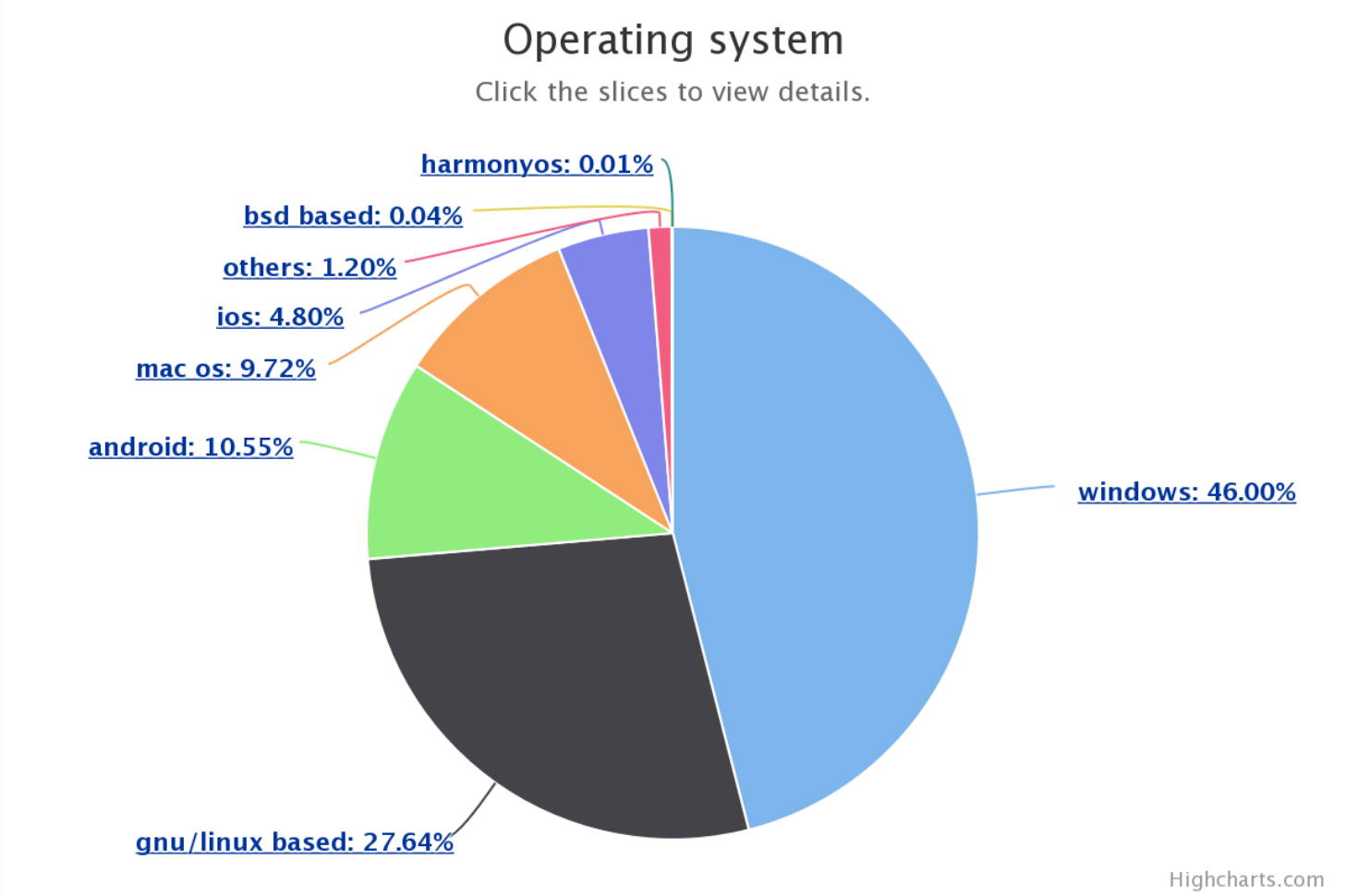

Figure 3: Répartition des operating systems sur 2.106.504 users, Le 09/10, de amiunique.org

Exemples : OS et navigateurs

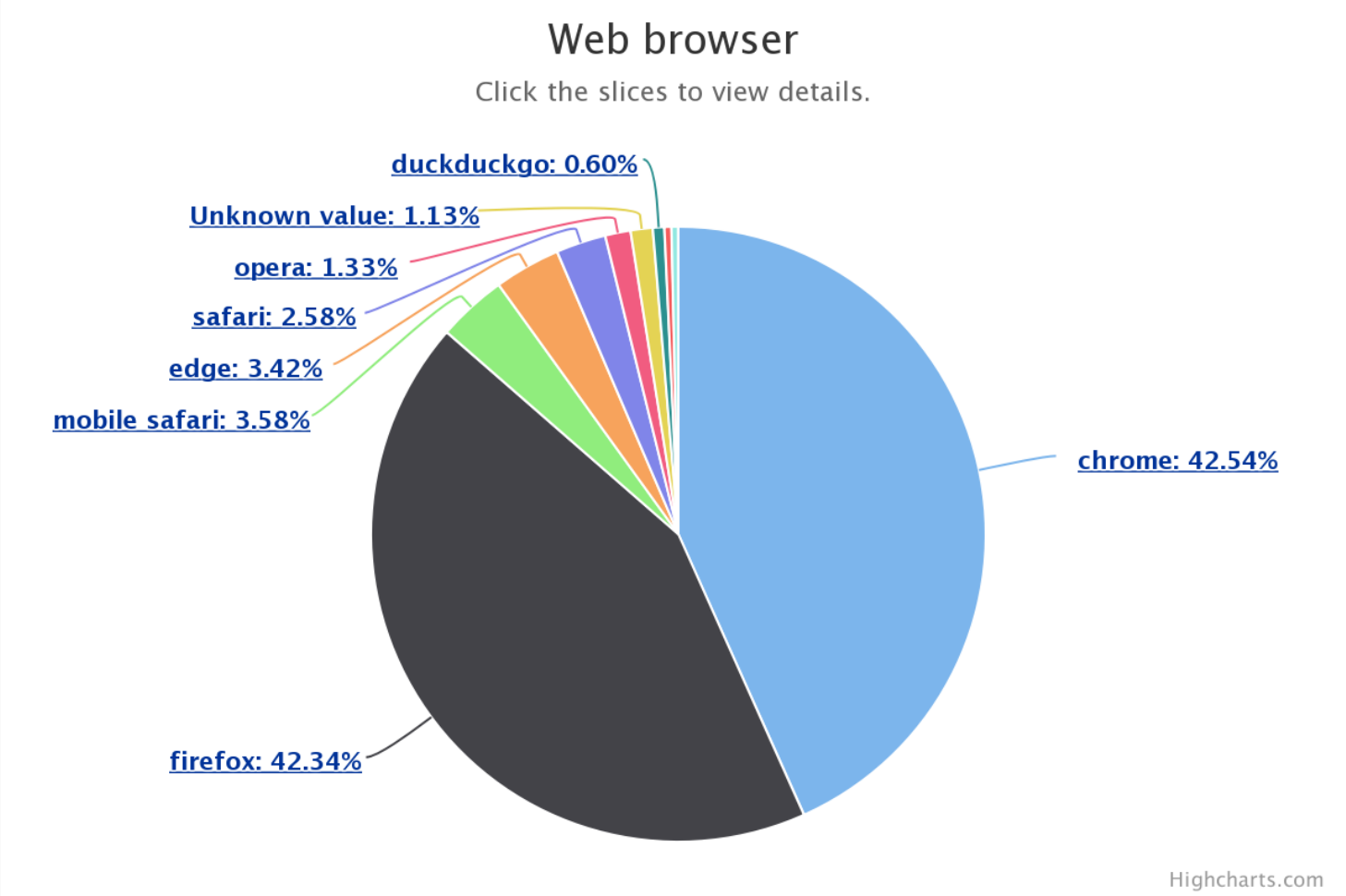

Figure 4: Répartition des navigateurs web sur 2.106.504 users, Le 09/10, de amiunique.org

1. Le principe du browser fingerprinting

1.2 Le browser fingerprinting: définition et utilisations

Le browser fingerprinting

Le browser fingerprinting

> Le fingerprinting, ou “prise d’empreinte” est une technique probabiliste visant à identifier un utilisateur de façon unique sur un site web ou une application mobile en utilisant les caractéristiques techniques de son navigateur.

Commission nationale de l’informatique et des libertés

Le browser fingerprinting

Le browser fingerprinting

> Le fingerprinting, ou "prise d'empreinte" est une technique **probabiliste** visant à identifier un utilisateur de façon unique sur un site web ou une application mobile en utilisant les caractéristiques techniques de son navigateur.

Commission nationale de l'informatique et des libertés

Attention aux préconçus

- Contrairement aux cookies, cela reste une méthode probabiliste: pas d'identification à coup sûr

Le browser fingerprinting

Le browser fingerprinting

> Le fingerprinting, ou "prise d'empreinte" est une technique **probabiliste** visant à identifier un utilisateur de façon unique sur un site web ou une application mobile en utilisant les **caractéristiques techniques** de son navigateur.

Commission nationale de l'informatique et des libertés

Attention aux préconçus

- Contrairement aux cookies, cela reste une méthode probabiliste: pas d'identification à coup sûr
- Le browser fingerprinting ne consiste pas qu'en de la collecte de données accessible en "clair" e.g. la taille de l'écran

Le browser fingerprinting

Le browser fingerprinting

> Le fingerprinting, ou "prise d'empreinte" est une technique **probabiliste** visant à identifier un utilisateur de façon unique sur un site web ou une application mobile en utilisant les **caractéristiques techniques** de son navigateur.

Commission nationale de l'informatique et des libertés

Le problème du consentement

- Le fingerprinting est une technique "stateless" : pas de stockage d'informations e.g. cookies

Le browser fingerprinting

Le browser fingerprinting

> Le fingerprinting, ou "prise d'empreinte" est une technique **probabiliste** visant à identifier un utilisateur de façon unique sur un site web ou une application mobile en utilisant les **caractéristiques techniques** de son navigateur.

Commission nationale de l'informatique et des libertés

Le problème du consentement

- Le fingerprinting est une technique "stateless" : pas de stockage d'informations e.g. cookies
- L'utilisateur ne peut ni accepter ni refuser le fingerprinting

Le browser fingerprinting

Le browser fingerprinting

> Le fingerprinting, ou "prise d'empreinte" est une technique **probabiliste** visant à identifier un utilisateur de façon unique sur un site web ou une application mobile en utilisant les **caractéristiques techniques** de son navigateur.

Commission nationale de l'informatique et des libertés

Une méthode dure à détecter

- Il dur de détecter si le fingerprinting est malicieux ou non
- Le fingerprinting est utile, e.g. pour détecter les bot (re-CAPTCHA)

Une méthode efficace ?

Que veut dire efficace ?

Il faut pouvoir identifier l'utilisateur de manière unique

Une méthode efficace ?

Que veut dire efficace ?

Il faut pouvoir identifier l'utilisateur de manière unique

Une efficacité importante

- En 2010, Eckersley a identifié 83% des ordinateurs uniquement
- En 2016, Laperdix et al. a identifié 90% des ordinateurs, et 80% des téléphones uniquement

Une méthode efficace ?

Que veut dire efficace ?

Il faut pouvoir identifier l'utilisateur de manière unique

Une efficacité importante

- En 2010, Eckersley a identifié 83% des ordinateurs uniquement
- En 2016, Laperdix et al. a identifié 90% des ordinateurs, et 80% des téléphones uniquement

Que veut dire efficace ?

Mais aussi de l'identifier durablement !

Une méthode efficace ?

Que veut dire efficace ?

Il faut pouvoir identifier l'utilisateur de manière unique

Une efficacité importante

- En 2010, Eckersley a identifié 83% des ordinateurs uniquement
- En 2016, Laperdix et al. a identifié 90% des ordinateurs, et 80% des téléphones uniquement

Que veut dire efficace ?

Mais aussi de l'identifier durablement !

Une efficacité importante

En 2018, Vastel et al. a montré qu'en moyenne un utilisateur est identifiable uniquement pendant 74 jours

Le browser fingerprinting, un accident ?

Le browser fingerprinting, un accident ?

C'est un accident !

Le fingerprint est un accident qui a été rendu possible par "l'amélioration" des navigateurs pour permettre plus d'interactions et de personalisation

Le browser fingerprinting, un accident ?

C'est un accident !

Le fingerprint est un accident qui a été rendu possible par "l'amélioration" des navigateurs pour permettre plus d'interactions et de personalisation

Deux technologies en cause

Les données sont majoritairement générées pour:

Le browser fingerprinting, un accident ?

C'est un accident !

Le fingerprint est un accident qui a été rendu possible par "l'amélioration" des navigateurs pour permettre plus d'interactions et de personalisation

Deux technologies en cause

Les données sont majoritairement générées pour:

- Avoir des architectures à plugins qui permettent de personnaliser l'apparence du navigateur, e.g. DarkReader

Le browser fingerprinting, un accident ?

C'est un accident !

Le fingerprint est un accident qui a été rendu possible par "l'amélioration" des navigateurs pour permettre plus d'interactions et de personalisation

Deux technologies en cause

Les données sont majoritairement générées pour:

- Avoir des architectures à plugins qui permettent de personnaliser l'apparence du navigateur, e.g. DarkReader
- Avoir API avancée qui permettent d'avoir des pages dynamique, e.g. dans la bonne language

API

API := Application Programming Interface

Une technique répandue

Rank Interval	Websites (count)	Websites (%)
1 to 1K	266	30.60%
1K to 10K	2,010	24.45%
10K to 20K	981	11.10%
20K to 50K	2,378	8.92%
50K to 100K	3,405	7.70%
1 to 100K	9,040	10.18%

Figure 5: Prévalence du FringerPrinting selon la popularité des site web (Alexa ranking) en 2020. fingerprinting the Fingerprinters: Learning to Detect browser fingerprinting Behaviors

Mais appliqué diversement

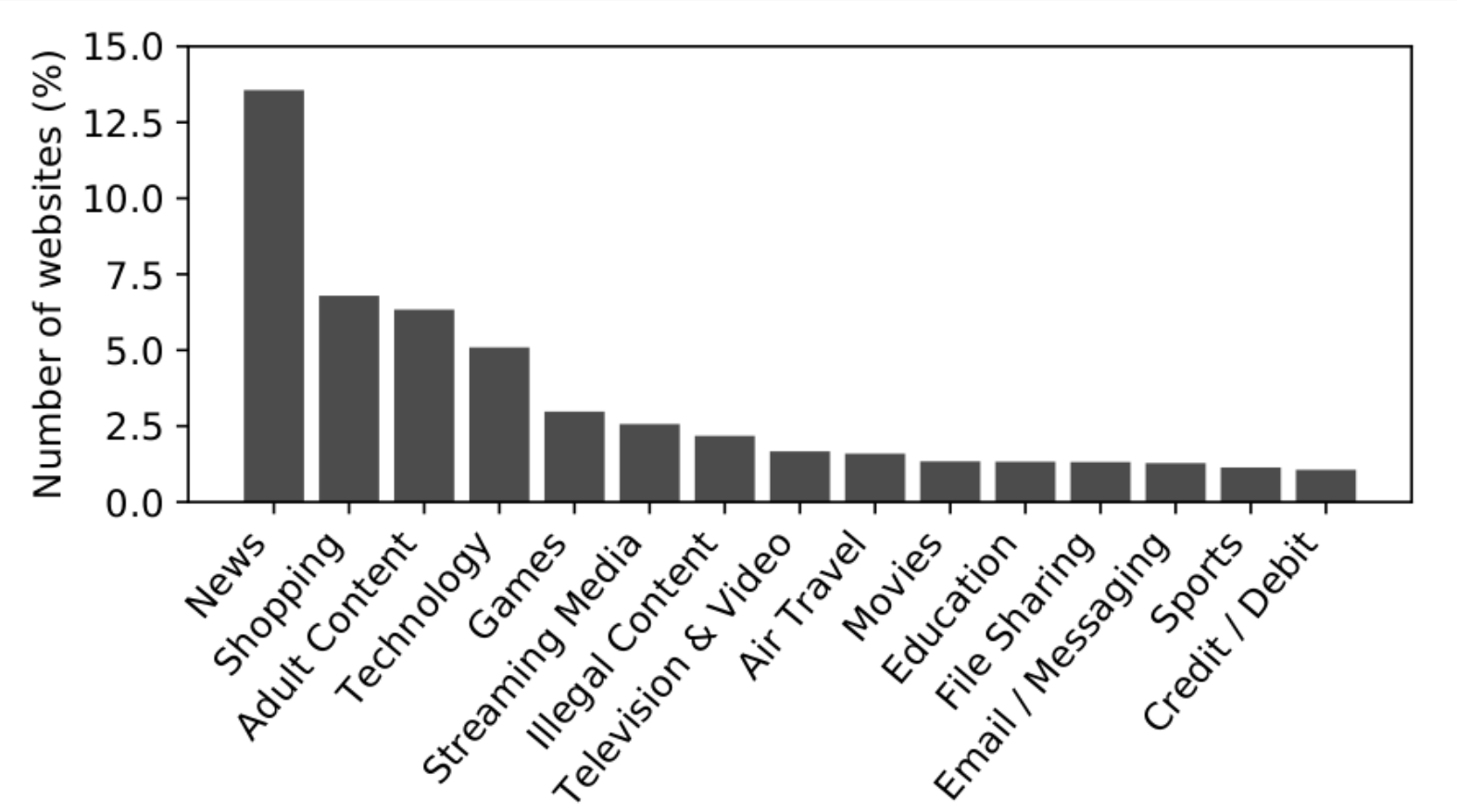

Figure 6: Prévalence du FringerPrinting selon le domaines des site web en 2020. fingerprinting the Fingerprinters: Learning to Detect browser fingerprinting Behaviors

2. Démonstration

Etes vous unique ?

Sur mon ordinateur, j'ai 20 identifieurs à moins de 5% sur 64 !
Sur mon iphone, j'ai 18 identifieurs à moins de 5% sur 64 !
Et vous ?

https://amiunique.org/fingerprint

3. Quelques méthodes de browser fingerprinting

3. Quelques méthodes de browser fingerprinting

3.2 La collecte de données par lecture des API

OS et Logiciels

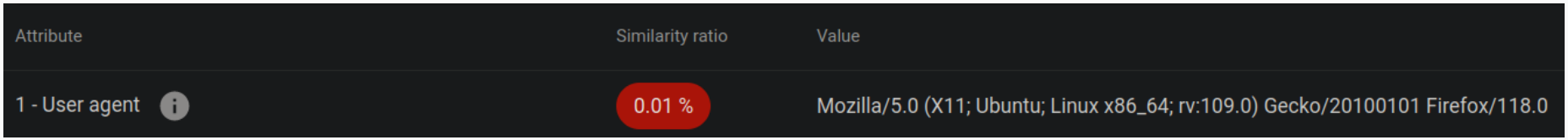

Figure 7: HTPP : User agent with Firefox

OS et Logiciels

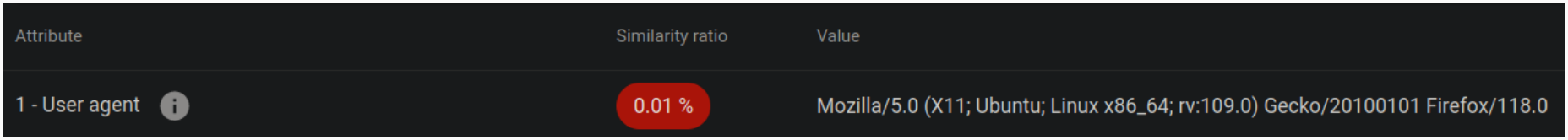

Figure 7: HTPP : User agent with Firefox

Des données discriminantes

Ces données sont discriminates si vous utilisez:

- Un OS peu utilisé, e.g. IOS 4.8%
- Un navigateur peu utilisé, e.g. edge 3.4%
- Un navigateur qui n'est pas à jour

OS, Logiciels et Browser

Attribute	Similarity ratio	Value
1 - User agent	0.01 %	Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:109.0) Gecko/20100101 Firefox/118.0

Figure 8: HTPP : User agent with Firefox

Attribut	Ratio de similarity	Valeur
1 - En-tête "User agent"	0.12 %	Mozilla/5.0 (iPhone; CPU iPhone OS 16_4_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/16.4 Mobile/15E148 Safari/604.1

Figure 9: HTPP : User agent with Safari

OS, Logiciels et Browser

Attribute	Similarity ratio	Value
1 - User agent	0.01 %	Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:109.0) Gecko/20100101 Firefox/118.0

Figure 8: HTPP : User agent with Firefox

Attribut	Ratio de similarity	Valeur
1 - En-tête "User agent"	0.12 %	Mozilla/5.0 (iPhone; CPU iPhone OS 16_4_1 like Mac OS X) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/16.4 Mobile/15E148 Safari/604.1

Figure 9: HTPP : User agent with Safari

Tous les navigateurs ne sont pas égaux

Certains navigateurs envoie plus de données que d'autres:

- Cela permet de discriminer selon la version précise de l'OS
- La version précise de l'ensemble de l'infrastructure navigateur

Exemple : Android

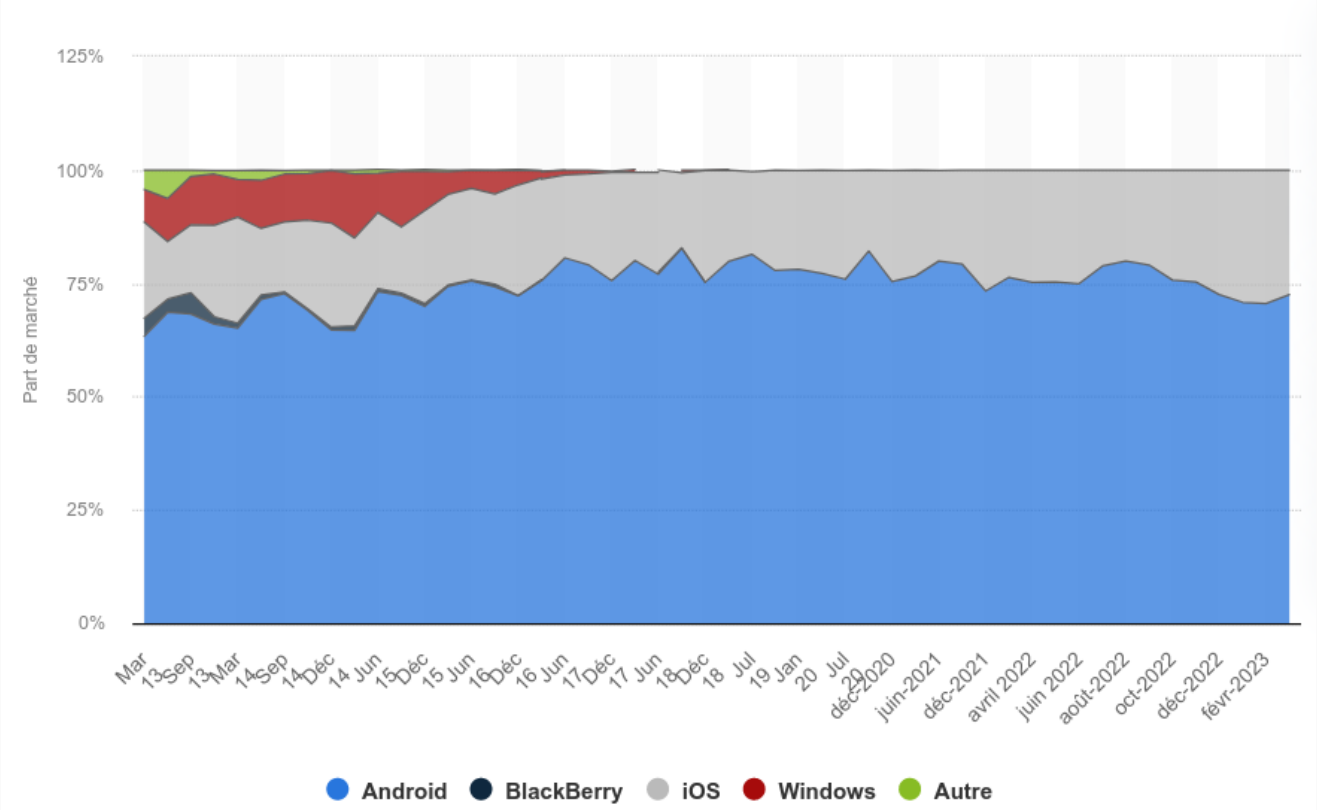

Figure 10: Part de marché sur smartphone,
Le 14/10, de Statistica

Exemple : Android

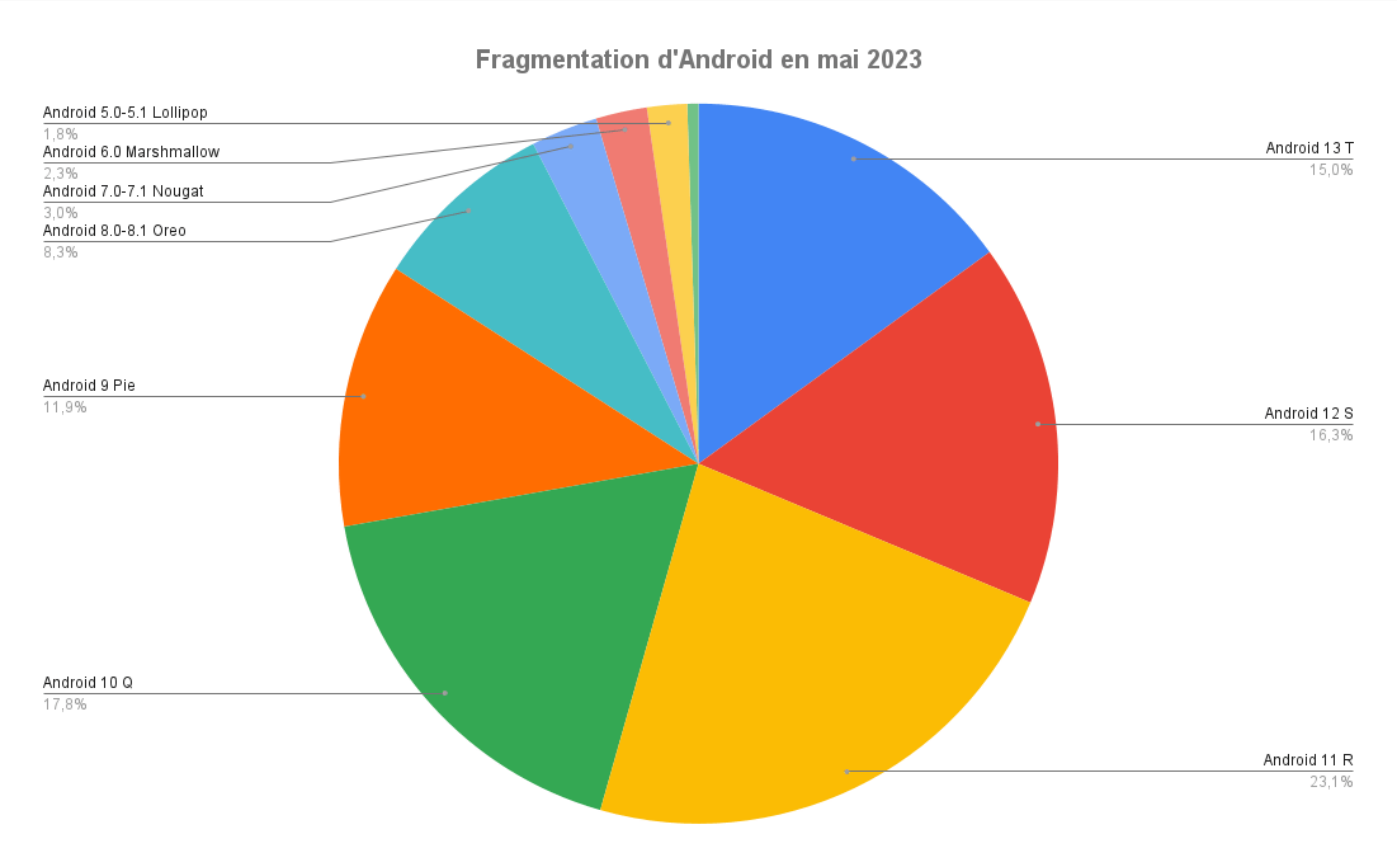

Figure 11: Part de marché sur smartphone,
Le 14/10, de Frandroid

Plugins

Plugins

Tous les navigateurs permettent d'ajouter des plugins qui permettent d'ajouter des fonctionnalités et de personnaliser l'apparence

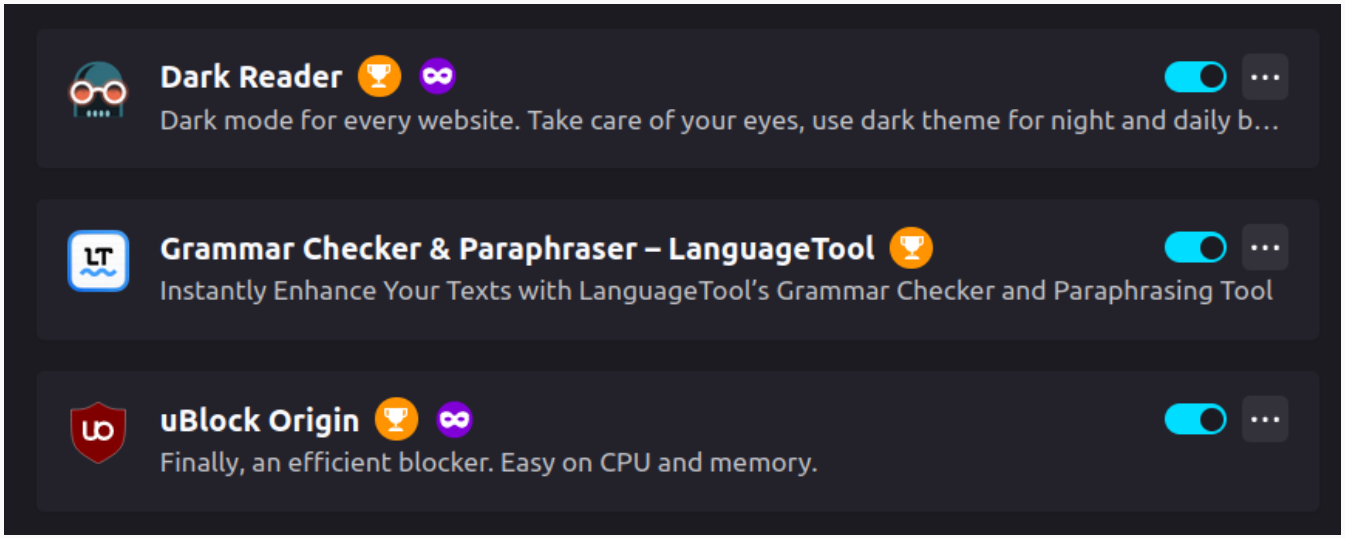

Figure 12: Exemples de plugins installés sur un ordinateur

Plugins

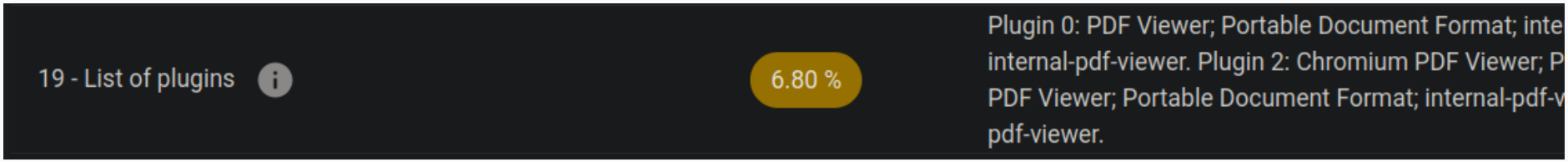

Figure 13: Javascript : list of plugins

Des données discriminantes

- Les plugins dépendent entièrement de qui vous êtes
- Il en existe un très grand nombre
- A part un minorité, ils sont peu répandus

Les polices d'écriture

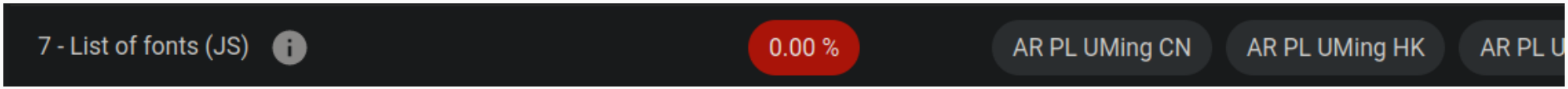

Figure 14: Javascript : partial list of fonts

Des données discriminantes

Les polices sont discriminates car:

- Les polices dependent de l'OS et des logiciels installé sur l'ordinateur
- Il existe un très grand nombre de polices d'écriture

D'autres données discriminantes

D'autres données discriminantes

- Récupérer les permissions du browser: notifications, caméra, ...
- Des informations sur l'affichage graphique avec WebGL
- Le layout du clavier: nombre de touches, touches spéciales (français ou suédois), disposition, ...
- Test de présence d'API, de la versions des API, ainsi que si des plugins les ont modifiés
- etc...

3. Quelques méthodes de browser fingerprinting

3.2 L'extraction de données par détournement des API

Le test du Canvas

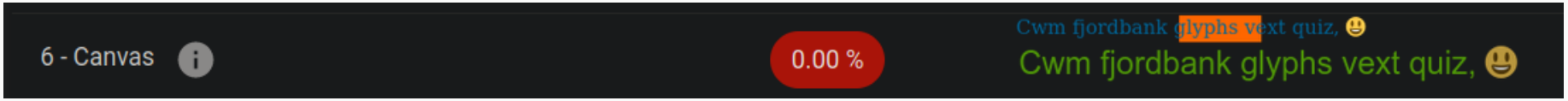

Figure 15: Javascript : détourner la fonction Canvas

Le test du Canvas

Il consiste à demander Canvas de Javascript de “dessiner” :

- Une phrase dans une police d’écriture qui n’existe pas
- Et d’afficher un emoji

Un test discriminant

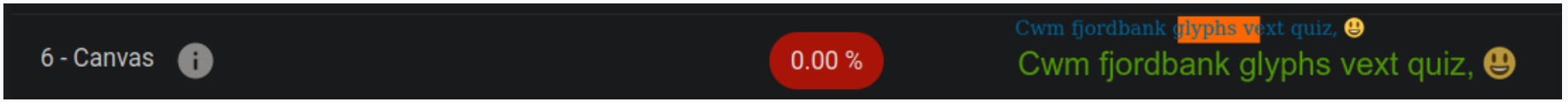

Figure 16: Javascript : détourner la fonction Canvas

Un test discriminant

- Si une police n'existe pas, le système utilise une "fallback" police qui dépend de l'OS mais aussi des polices installées

Un test discriminant

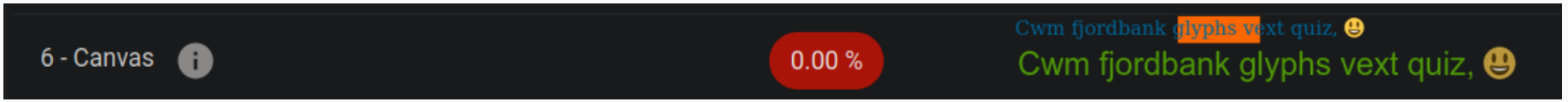

Figure 16: Javascript : détourner la fonction Canvas

Un test discriminant

- Si une police n'existe pas, le système utilise une "fallback" police qui dépend de l'OS mais aussi des polices installées
- Les emojis ne sont pas standard mais fournit par l'OS

Un test discriminant

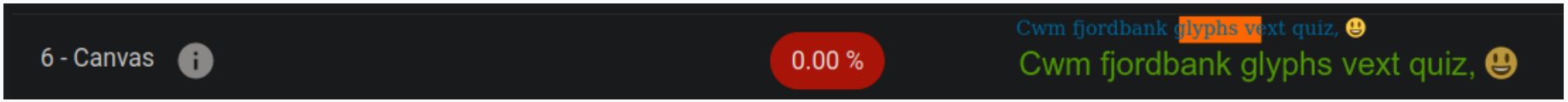

Figure 16: Javascript : détourner la fonction Canvas

Un test discriminant

- Si une police n'existe pas, le système utilise une "fallback" police qui dépend de l'OS mais aussi des polices installées
- Les emojis ne sont pas standard mais fournit par l'OS
- L'affichage dépend aussi du hardware, dont la carte graphique

D'autres données discriminantes

D'autres données discriminantes

- Animer une image pour calculer le nombre d'images par seconde
- Faire une requête pour calculer le temps de réponse du DNS, et déterminer si une page a été consulté récemment
- Diffuser des sons pour calculer le profil audio de l'ordinateur
- etc...

4. Des solutions de contournement ?

Les défis du browser fingerprinting

Quelques défis

Il est dur de contourner le fingerprinting car :

- Il est dur de déterminer si un script est malveillant ou non
- Bloquer un script peut casser une page e.g. script de connection
- Bloquer des API e.g. Canvas peut casser une page
- Certains scripts entremêlent fingerprinting et fonctionnalités classiques

Modifier les données

Modifier les données

Randomizer : Ajouter un bruit aléatoire aux données renvoyées

Standardizer : Renvoyer les même données pour tout le monde

Modifier les données

Modifier les données

Randomizer : Ajouter un bruit aléatoire aux données renvoyées

Standardizer : Renvoyer les même données pour tout le monde

Limitations

- Ces deux méthodes peuvent impacter l'expérience utilisateur
- L'ajout de bruit peut être éliminé
- La standardization peut faire croire à des bots
- Ces méthodes peuvent aider à identifier l'utilisateur

Bloquer des fonctionnalités

Bloquer des fonctionnalités

- Bloquer le téléchargement de script de browser fingerprinting
- Bloquer certaines API comme Canvas ou WebGL...

Bloquer des fonctionnalités

Bloquer des fonctionnalités

- Bloquer le téléchargement de script de browser fingerprinting
- Bloquer certaines API comme Canvas ou WebGL...

Limitations

Cela peut tout casser

Bloquer des fonctionnalités

Bloquer des fonctionnalités

- Bloquer le téléchargement de script de browser fingerprinting
- Bloquer certaines API comme Canvas ou WebGL...

Limitations

Cela peut tout casser

Policy	Major (%)	Minor (%)	Total (%)
Blanket API restriction	48.36%	19.67%	68.03%
Targeted API restriction	24.59%	5.73%	30.32%
Request blocking	44.26%	5.73%	50%
Hybrid	38.52%	8.19%	46.72%

Figure 17: Study case: Cassage des sites web selon la méthode de blocage

Conclusion

Conclusion

todo