



Mémoire de Stage de 4e année

# FL-Minifer

## A Tool To Minify and Unify AdBlocker's Filter Lists

Maxence NEUS



### Polytech Lille

Secrétariat - Bureau  
Boulevard Paul Langevin - Cité  
Scientifique  
59655 VILLENEUVE D'ASCQ  
CEDEX  
03-28-76-73-60  
03-28-76-73-61



### Inria

40 Av. Halley  
59650 Villeneuve d'Ascq

Tuteur Entreprise :

**Walter RUDAMETKIN**

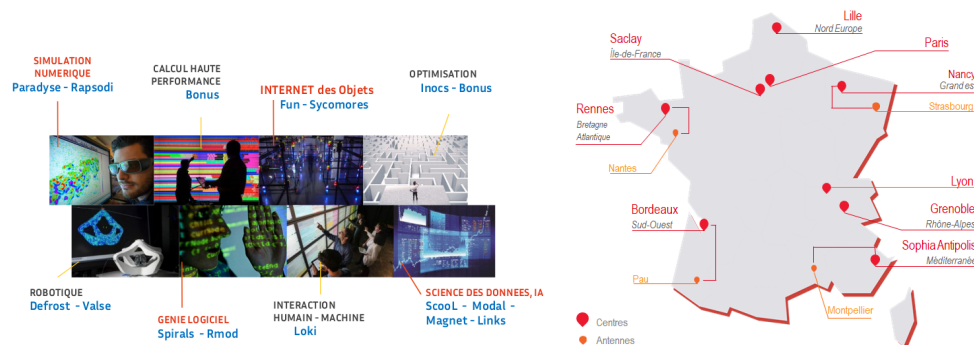
Tuteur Ecole

**Walter RUDAMETKIN**

2022

# Contents

<b>1</b>	<b>Présentation de l'entreprise</b>	<b>2</b>
1.1	Inria . . . . .	2
1.2	SPIRALS . . . . .	3
<b>2</b>	<b>Contexte du projet</b>	<b>4</b>
2.1	Introduction aux adBlockers . . . . .	4
2.2	Présentation AmIUnique . . . . .	5
<b>3</b>	<b>Corps du projet</b>	<b>5</b>
<b>4</b>	<b>Extras</b>	<b>5</b>



# 1 Présentation de l'entreprise

## 1.1 Inria

Le centre de recherche Inria fondé en 1967 est un établissement public de recherche se spécialisant dans le domaine des mathématiques et de l'informatique. Le centre de Lille Nord Europe fait partie des 9 centres autour de la France et comporte deux complexes: à la Haute Borne et à Euratechnologie. Au sein du centre dirigé par Mireille RÉGNIER opèrent 15 équipes dont mon équipe d'accueil: SPIRALS.

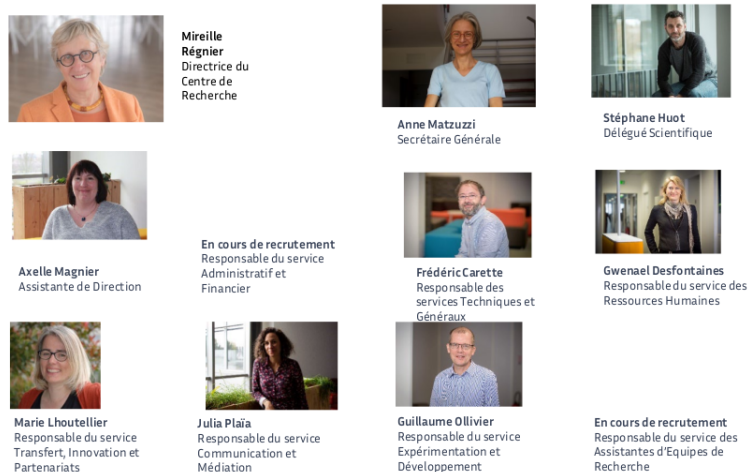


Figure 1: Équipe de direction

## **1.2 SPIRALS**

est une équipe jointe de l'INRIA et de CRISTAL qui se focalise sur les systèmes distribués et l'ingénierie système. les projets de l'équipe comportent l'étude de solutions autonomes, efficaces et adaptatives pour la récupération et le traitement de données. Dirigée par Lionel Seinturier, l'équipe est actuellement composée de 11 chercheurs permanents, 7 postdocs, 17 doctorants et 7 ingénieurs.

## 2 Contexte du projet

### 2.1 Introduction aux adBlockers

**Les cookies:** Le modèle économique du web est centré autour des données utilisateur et de la façon dont elles sont utilisées, majoritairement pour cibler au mieux des publicités qui pourront alors faire un maximum de conversions (nombres d’achats par publicité imprimée sur une page). Evidemment pour cibler efficacement un utilisateur il est nécessaire de pouvoir associer un ensemble de données à un utilisateur, historiquement les cookies ont été l’outil le plus répandu pour remplir cette fonction: en incorporant un identifiant unique à chaque utilisateur dans ses requêtes HTML.

Malheureusement pour les géants du web, l’idée que quelques entreprises puissent avoir accès à une quantité alarmante de leurs données simplement par leur navigation sur le web ne plaît pas vraiment aux utilisateurs. Suite à des campagnes de sensibilisation de plus en plus d’utilisateurs du web se sont retrouvés confrontés à cette réalité et désirent autant que possible limiter l’accès à leurs données de navigation. Pour cela plusieurs compagnies se sont mises à créer des extensions pour navigateurs qui permettent de bloquer les dis “tracking cookies”. En effet tous les cookies n’ont pas pour but de traquer l’utilisateur, la technologie a été créée pour garder des informations utilisateurs essentielles à la fonctionnalité du site comme rester connecter à son compte utilisateur ou sauvegarder l’état du panier sur les sites de shopping en ligne. Ces cookies dits “essentiels” doivent être filtrés pour garder les fonctionnalités du site, une méthode largement utilisée est de bloquer les cookies dits “tierces” qui proviennent de domaines différents de celui du site visité étant donné que ceux ci sont bien souvent utilisés uniquement comme tracking cookies et peuvent donc être bloqués sans crainte de nuire aux fonctionnalités du site.

**La publicité sur le web:** Comme dit précédemment, le modèle économique du web consiste à montrer le plus de publicités intéressantes possible à l’utilisateur pour qu’il achète le plus de produits possible. Ce modèle entraîne logiquement un comportement où les sites utilisent chaque pixel d’espace libre de contenu sur leurs sites pour y servir des publicités grâce à un modèle d’enchères où les distributeurs de publicités enchérissent le plus sur les utilisateurs avec un intérêt probable pour leurs produits.

Les utilisateurs veulent donc se débarrasser un maximum des publicités à la fois pour leur confort mais aussi pour protéger les plus vulnérables des pratiques prédatrices des entreprises. C’est là l’utilité des adBlockers qui

scannent les éléments de la page et bloquent ceux qui sont détectés comme étant des publicités.

**Les règles de filtrage:** Pour détecter les éléments qui font parties de publicités, les adBlockers utilisent des listes de filtres qui contiennent des caractéristiques d'éléments publicitaires rencontrés sur le web. Ces listes sont maintenues régulièrement par la communauté et par des entreprises d'adBlockers pour parer aux nouvelles tentatives de contournement des distributeurs de publicités. Les listes sont de simples fichiers texte dont la syntaxe de base est définie par AdBlockPlus (ABP) [ici](#) et elle est étendue par uBlockOrigin (uBO) [ici](#). La syntaxe comporte deux grands types de règles:

- Les *Network Rules*
- Les *Cosmetic Rules*

Le cas le plus simple, les *Network Rules* correspondent à des règles qui bloquent les éléments selon l'url dont ils sont importés, ce grâce à une syntaxe similaire à des expressions régulières. C'est à dire que la règle décrits un modèle d'url qui correspond à un ensemble d'url qui devront être bloquées, par exemple "http://example.com/ads/banner\*.gif" bloquera les éléments venant de la source "http://example.com/ads/banner420.gif" (le caractère \* correspondant à n'importe quelle suite de caractère) mais pas "http://example.fr/ads/ad123.png".

Plus puissantes, les *Cosmetic Rules* bloquent des éléments dont le css correspond à une certaine description, par exemple la règle "###ad-boxes" bloque les éléments sur la page qui ont une id css de "ad-boxes".

Il est également possible d'écrire des exceptions qui permettent de laisser passer des éléments qui seraient bloqués par une règle trop générale et qui nuirait aux fonctionnalités du site.

## 2.2 Présentation AmIUnique

Le Fingerprinting sur le web [1]

## 3 Corps du projet

## 4 Extras

## References

- [1] Benoit Baudry Pierre Laperdrix Walter Rudametkin. “Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints”. In: *IEEE Symposium on Security and Privacy* (2016).

## **Résumé**

Lors de ce stage blablabla

## **Abstract**

During this internship blablabla