Défense de mémoire

Analyse du traçage des utilisateurs via leur navigateur Web

Antoine Marchal

23 juin 2014

Plan

Introduction

- Intensification de la surveillance sur Internet
- Le respect de la vie privée est de plus en plus mis à mal

Principes importants

- Protocole HTTP
- Cookies
- Cache
- Principe de même origine

Moyens d'identification

- Cookies
- Cache
- Pixels espions
- JavaScript
- Flash
- Empreintes des navigateurs

Outil implémenté

- Crawler : visite les sites et enregistre des informations (requêtes et réponses HTTP)
- Parser : traite les informations enregistrées afin d'identifier des trackers

Critères d'enregistrement d'informations

- Présence de l'URL dans la base de données Ghostery

- La ressource chargée est du JavaScript
- La ressource chargée est du JavaScript et l'URL contient des paramètres (chaîne de requête)
- La ressource chargée est du Flash
- Les images chargées ont une largeur et hauteur de 1 pixel
- Un cookie est créé via la réponse HTTP
- L'URL des ressources chargées contient des paramètres

Discussion sur les critères

- Influence de l'utilisation de la base de données Ghostery
- Influence de l'ordre des critères d'identification

Résultats

2 types d'analyse :

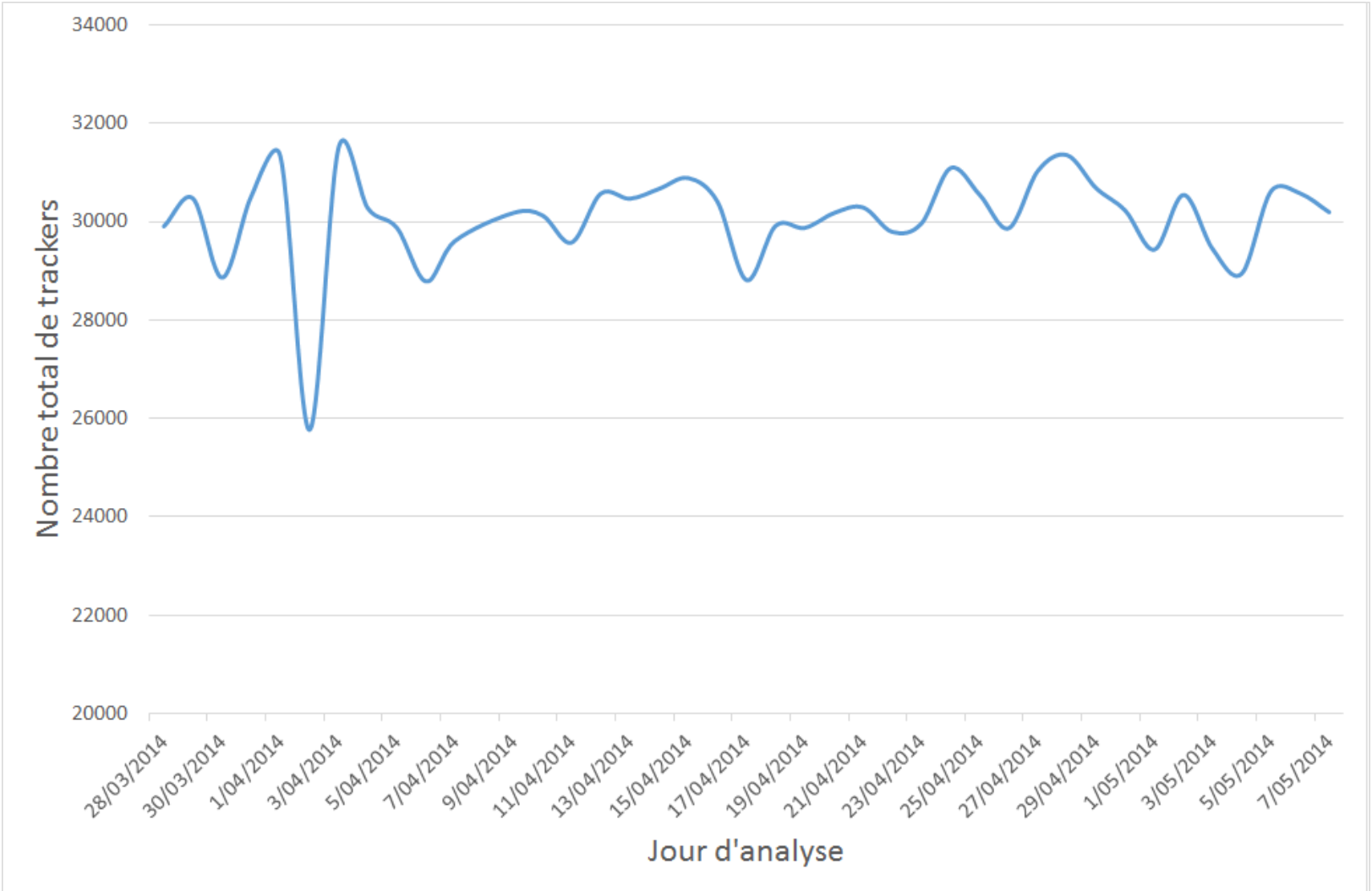
- Analyse à long terme : exécution régulière du *crawler*
  Expérience 1 : 40 jours d'analyses, visite quotidienne

- Analyse à court terme : exécution ponctuelle du *crawler*
  Expérience 2 : visite ponctuelle

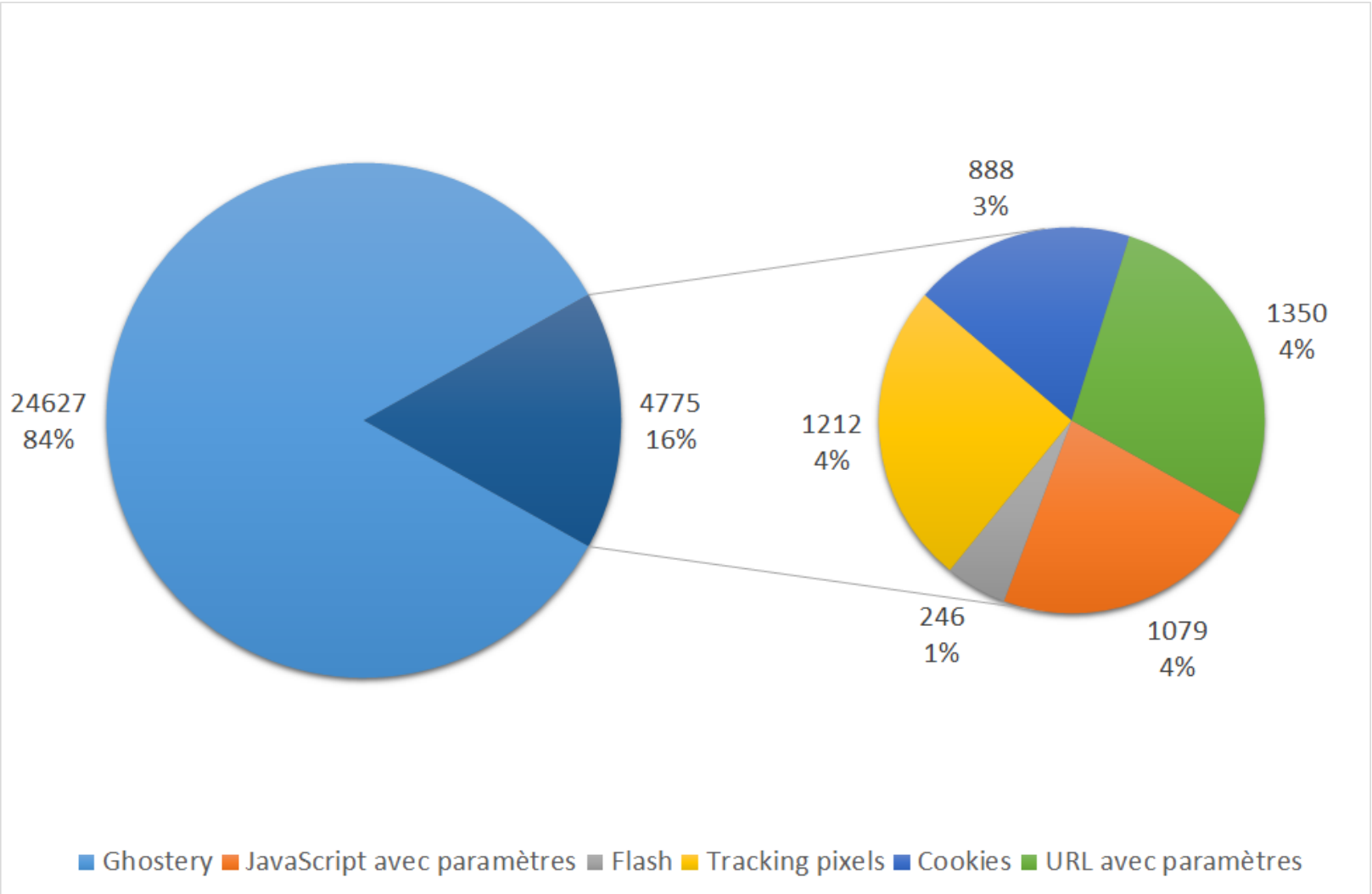
Analyse à long terme

- Stabilité du *crawler* : $< 9\%$ d'échec
  (analyse de 1000 sites - classement du TOP Alexa)
- Chute du taux de réussite du *parser*
- Nombre de trackers détectés reste relativement stable

Evolution du nombre de trackers



Analyse à court terme



Organisations déployant le plus de trackers

Sur un total de 24627 trackers :

- 2562 : DoubleClick
- 1379 : AppNexus
- 949 : Google Analytics
- 675 : Google Adsense
- 574 : Rubicon
- 534 : Turn
- 441 : ScoreCard Research Beacon
- 347 : MediaMath
- 332 : Lotame
- 326 : OpenX

Défenses envisageables

- Extensions de navigateur testées
  (Adblock Plus, DoNotTrackMe, Ghostery, HTTPSEverywhere, Priv3, Privacy Badger,...)
- Do Not Track
- Autres extensions intéressantes (BetterPrivacy, NoScript,...)

Comparaison des défenses testées

Exécution du *parser* avec utilisation de la base de données Ghostery

	Ghostery	JS avec paramètres	Flash	Tracking pixels	Cookies	URL avec param.	TOTAL
Normal (sans extension)	24627	1079	246	1212	888	1350	29402
Adblock (par défaut)	2584	558	118	89	303	563	4215
Adblock (pas de publicité acceptable)	2572	544	134	100	207	570	4127
Adblock (Fanboy Ultimate)	2197	497	139	99	168	538	3638
DoNotTrackMe	9341	812	227	424	567	852	12223
Ghostery	471	553	130	200	319	551	2224
HTTPS Everywhere	19388	997	210	1017	821	1214	23647
Priv3	21670	963	230	1139	869	1142	26013
Privacy Badger	16215	1027	256	661	871	1069	20099
Adblock & Ghostery	448	509	123	96	258	471	1905
DNT	22533	1068	261	1175	855	1378	27270

Exécution du *parser* sans utilisation de la base de données Ghostery

	JS avec paramètres	Flash	Tracking pixels	Cookies	URL avec param.	TOTAL
Normal (sans extension)	4966	670	7809	8690	4788	26923
Adblock (par défaut)	924	131	169	348	812	2384
Adblock (pas de publicité acceptable)	907	152	171	274	814	2318
Adblock (Fanboy Ultimate)	664	155	168	218	772	1977
DoNotTrackMe	2290	358	1951	3296	1525	9420
Ghostery	558	130	210	328	558	1784
HTTPS Everywhere	4016	485	5926	6406	3755	20588
Priv3	4472	574	6382	8076	3811	23315
Privacy Badger	3793	411	4478	5571	2717	16970
Adblock & Ghostery	512	123	101	260	474	1470
DNT	4701	688	6798	7072	4604	23863

Constatations

- Les meilleures extensions sont celles qui reposent sur une base de données de trackers
- Une extension (Privacy Badger) qui analyse le comportement des sites mérite une attention particulière (version stable)
- L'entête HTTP Do Not Track permet une baisse du nombre de trackers : signe encourageant

Quelques améliorations possibles

- Meilleure intégration du module de visite des sites
  - Gestion des erreurs plus efficace
  - Analyse du code source

- Analyse du comportement des codes JavaScript et Flash

Conclusion

- La vie privée sur Internet est un sujet d'actualité
- Vue d'ensemble des techniques de traçage utilisées
- Base solide pour toute personne s'intéressant à la vie privée

Merci de votre attention.