

Vie privée et nouveaux modes de consommation de contenu culturel

Mathieu Cunche

INSA-Lyon CITI, Inria Privatics



Université Populaire, CAPI, 22/11/2022

Mathieu Cunche :

- Maître de conférence à l'INSA-Lyon en informatique
 - Membre de l'équipe Inria Privatics
- Enseignement en informatique : programmation, algorithmique, sécurité
- Recherches sur la **protection de la vie privée**
 - Réseaux sans-fils, objets connectés, applications mobiles ...
 - Identification de failles et conception de mécanismes de protection



- 1 Introduction
 - Culture et vie privée
 - Numérisation de l'offre culturelle
 - Collecte et exploitation des données
- 2 "Anonymisation" de profils Netflix
- 3 Applications mobiles de streaming
- 4 Le cas des TV connectées
- 5 Conclusion

Consommation de contenu culturel et vie privée

- Consommation essentiellement dans l'espace privé¹
- Pratique souvent personnelle
- Reflétant des traits personnels
 - Centre d'intérêts ...



¹Pasquier, "L'espace privé comme lieu de consommation culturelle".

Dématérialisation des contenus

- Plateformes de streaming video et audio
- Liseuses, lecteurs MP3
- Applications mobiles
- TV connectés



Le numérique et les données

- Ce que permet le numérique
 - Communication "instantanée", automatisation des tâches, dématérialisation ...
 - Collecte de données ...

²<https://whotracks.me/trackers.html>

Le numérique et les données

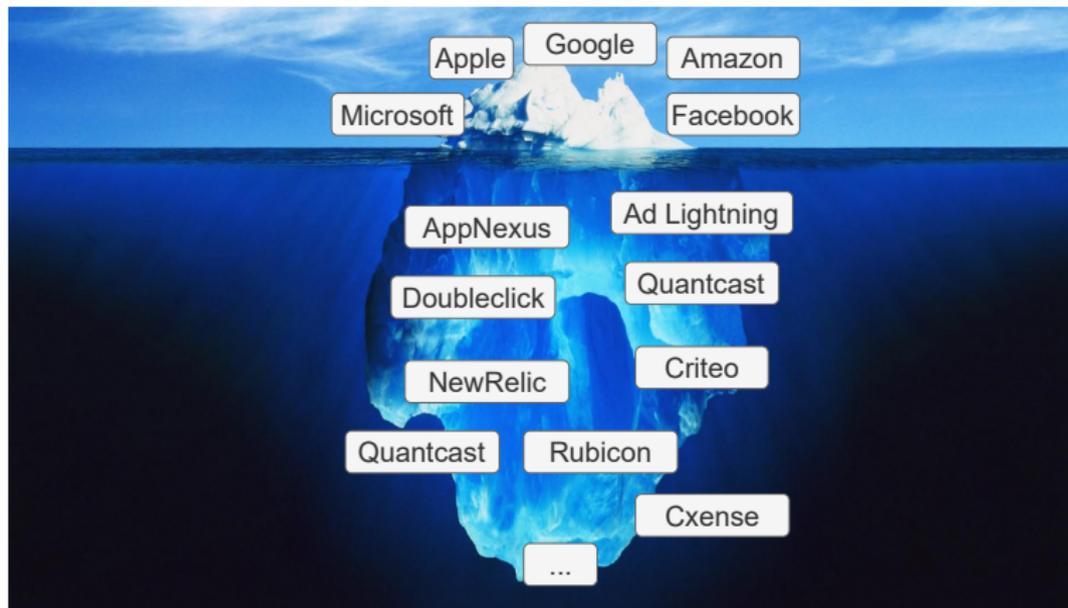
- Ce que permet le numérique
 - Communication "instantanée", automatisation des tâches, dématérialisation ...
 - Collecte de données ...
- Profilage des personnes
 - Ciblage : publicité personnalisée
 - Adaptation des offres : évaluation du risque (banques, assurances, ...)
 - Surveillance (policière, militaire, ...)



2

²<https://whotracks.me/trackers.html>

Exploitation des données : un écosystème complexe



- 1 Introduction
 - Culture et vie privée
 - Numérisation de l'offre culturelle
 - Collecte et exploitation des données
- 2 "Anonymisation" de profils Netflix
- 3 Applications mobiles de streaming
- 4 Le cas des TV connectées
- 5 Conclusion



- Compétition pour des algorithmes de recommandation (Netflix 2006)
 - Jeu de données avec des profils d'utilisateurs "anonymisés"
 - Objectif : concevoir un algorithme de prédiction des appréciations des films

	Ben-Hur	Matrix	Léon	Snatch	...
u_1	***		****	**	
u_2		***	**		
u_3	**		*	***	
...					
u_k	***	*	**	?	

- Pseudonymisation : l'identité des utilisateurs a été retirée du jeu de données
 - Impossible de réidentifier **directement** les utilisateurs

Attaque de réidentification I



- Une source auxiliaire d'information : IMDB
 - IMDB : *Internet Movie Database*
 - Site mondial répertoriant des données sur le cinéma
 - Films notés par les utilisateurs (55 millions)
 - Un utilisateur peut donner une note aux films qu'il a visionné
 - Déclaration facultative : un utilisateur peut décider de ne pas noter certains films qu'il aurait visionné

Attaque de réidentification II

- Des chercheurs démontrent que le jeu de données n'est pas anonyme³
 - Il est possible de réidentifier les utilisateurs
 - En combinant le dataset avec les profils publics d'IMDB
- Méthode de ré-identification basée sur la similarité
 - Etant donné un profil Netflix X trouver le profil IMDB Y qui lui ressemble le plus



Attaque de réidentification III

- Profil d'utilisateur : une liste d'éléments

$$X = (\text{Ben-Hur}, \text{Léon}, \text{Snatch})$$

- Exemple de calcul de score de similarité : l'index de Jaccard

- Formule :

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

- Le nombre d'éléments commun aux deux profils ($|X \cap Y|$) divisé par le nombre total d'éléments dans les profils ($|X \cup Y|$)
 - Similarité $\rightarrow 0$: profils différents
 - Similarité $\rightarrow 1$: profils similaires
- Application :

	Ben-Hur	Matrix	Léon	Snatch
X	✓		✓	✓
Y	✓		✓	
Z		✓		

- $J(X, Y) = 2/3 = 0.666$ et $J(X, Z) = 0/4 = 0$

Attaque de réidentification IV

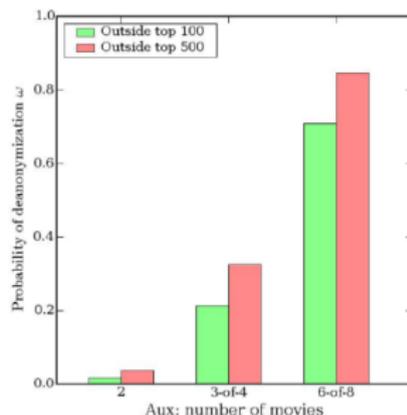


Figure 8. Adversary knows exact ratings but does not know dates at all.

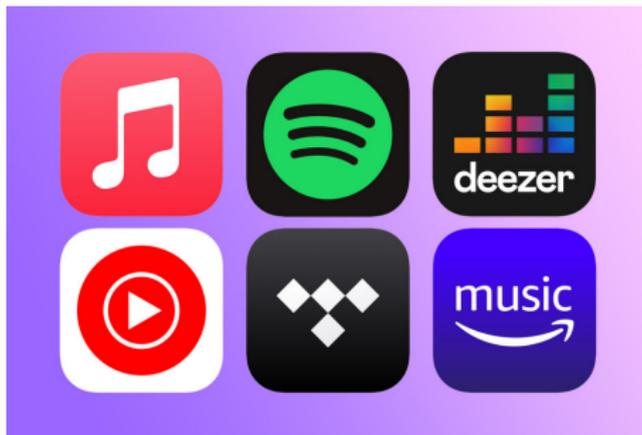
Figure: Fraction d'utilisateurs réidentifiés en fonction du nombre de films dans leur profil Netflix.

- Environ 80% des utilisateurs avec 6-8 films dans leur profil sont réidentifiables

³Narayanan and Shmatikov, "Robust De-anonymization of Large Sparse Datasets".

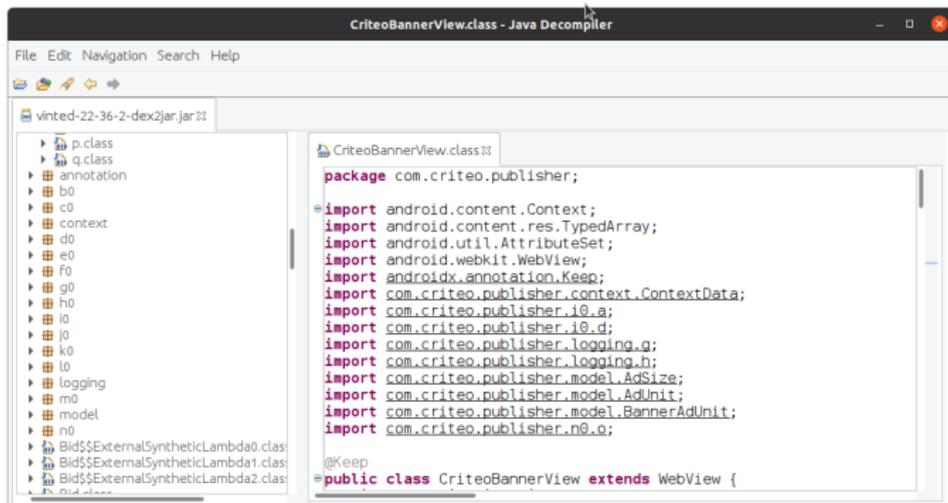
- 1 Introduction
 - Culture et vie privée
 - Numérisation de l'offre culturelle
 - Collecte et exploitation des données
- 2 "Anonymisation" de profils Netflix
- 3 Applications mobiles de streaming
- 4 Le cas des TV connectées
- 5 Conclusion

Applications mobiles de streaming



Méthodologie : comment analyser un app 1

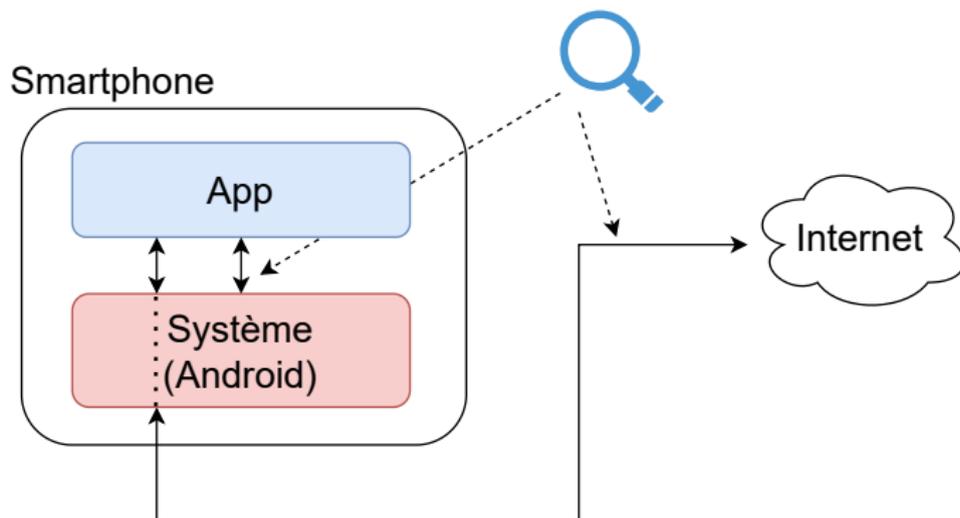
- Analyse statique : étude du **code** de l'application
 - Identification des permissions demandées, des trackers, des fonctionnalités potentiellement utilisées ...



```
CriteoBannerView.class - Java Decompiler
File Edit Navigation Search Help
vinted-22-36-2-dex2jar.jar
  p.class
  q.class
  annotation
  b0
  c0
  context
  d0
  e0
  f0
  g0
  h0
  i0
  j0
  k0
  l0
  logging
  m0
  model
  n0
  Bid$$ExternalSyntheticLambda0.class
  Bid$$ExternalSyntheticLambda1.class
  Bid$$ExternalSyntheticLambda2.class
  Bid.class
CriteoBannerView.class
package com.criteo.publisher;
import android.content.Context;
import android.content.res.TypedArray;
import android.util.AttributeSet;
import android.webkit.WebView;
import androidx.annotation.Keep;
import com.criteo.publisher.context.ContextData;
import com.criteo.publisher.i0.a;
import com.criteo.publisher.i0.d;
import com.criteo.publisher.logging.g;
import com.criteo.publisher.logging.h;
import com.criteo.publisher.model.AdSize;
import com.criteo.publisher.model.AdUnit;
import com.criteo.publisher.model.BannerAdUnit;
import com.criteo.publisher.n0.g;
@Keep
public class CriteoBannerView extends WebView {
```

Méthodologie : comment analyser un app II

- Analyse dynamique : étude du comportement de l'application
 - Observation de l'application pendant son utilisation
 - Données personnelles obtenues, transmises (à qui ?), ...





- Politique de confidentialité de l'app : <https://www.deezer.com/legal/personal-datas>
- Rapport Exodus : <https://reports.exodus-privacy.eu.org/fr/reports/deezer.android.app/latest/>

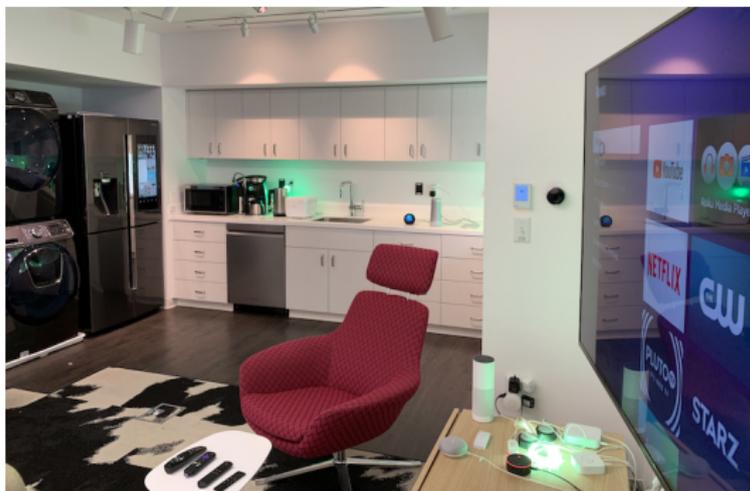
- 1 Introduction
 - Culture et vie privée
 - Numérisation de l'offre culturelle
 - Collecte et exploitation des données
- 2 "Anonymisation" de profils Netflix
- 3 Applications mobiles de streaming
- 4 Le cas des TV connectées
- 5 Conclusion



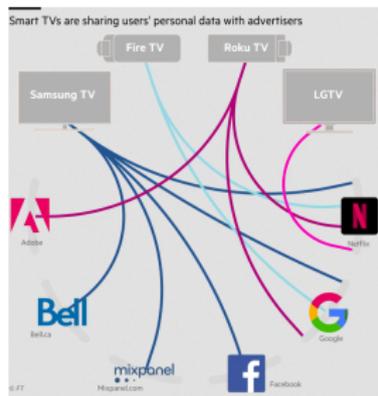
- Un écran enrichi avec des fonctionnalités
 - Streaming, Navigation Web, Jeux, Réseaux sociaux, ...
- Connecté à Internet

Données collectées par les TV I

- Analyse du trafic réseau généré par des TV – et autres objets – connectés⁴
 - TV installées dans laboratoire équipé de moyens d'écoute



Données collectées par les TV II

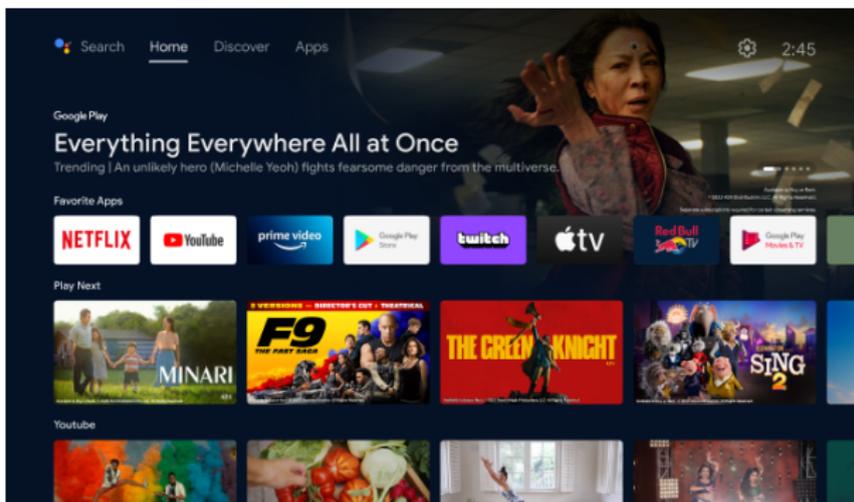


- Les TV connectés communiquent régulièrement avec des serveurs distants
 - Fabricant (ex.: Samsung) et parties tierces (Microsoft, Facebook, ...)
- Des échanges de données nombreux et pas toujours justifiés
 - Communication avec Netflix, même si aucun compte Netflix n'a été configuré
- Informations transmises
 - Geolocalisation, modèle de TV, activité ...

⁴Ren et al., "Information Exposure From Consumer IoT Devices".

Applications pour TV connectées I

- Comme les smartphones, les smart TV peuvent aussi accueillir des applications
 - Par exemple avec le système Android TV



Applications pour TV connectées II

- Etude des applications Android TV⁵
 - Collecte de données personnelles : identité, localisation, ...

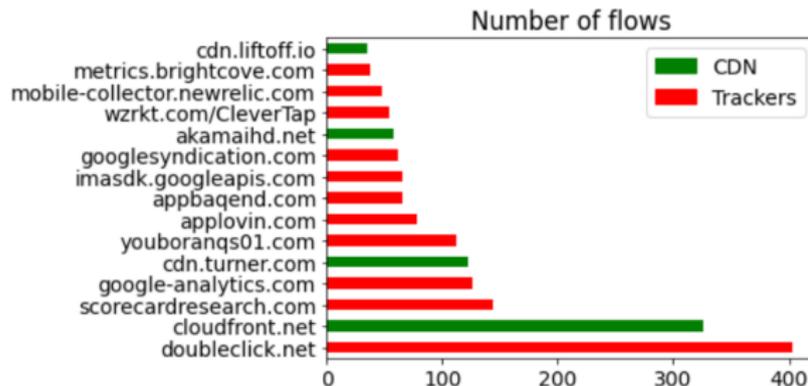
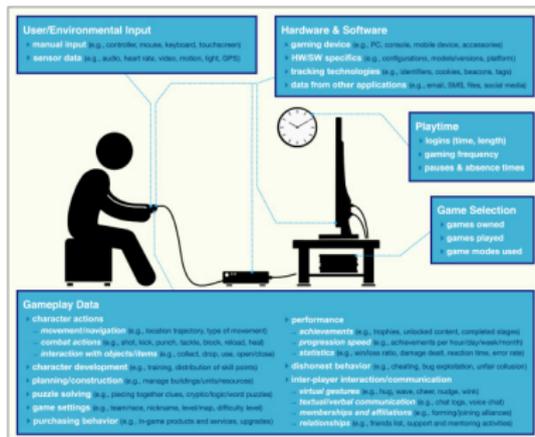


Fig. 5. Top domain names collecting sensitive data.

- Parmi les 15 destinations les plus fréquemment contactées par ces apps
 - 3 sont des fournisseurs de contenu (CDN)
 - les autres sont des traqueurs bien connus ...

⁵Tileria and Blasco, "Watch Over Your TV".

Exemple des jeux vidéos 1



- Informations pouvant être inférées⁶
 - Identité (nom, adresse email, réseaux sociaux), age, genre, émotions, compétence (résolution de problèmes, modélisation spatiale, mémorisation, ...), centres d'intérêt, pouvoir d'achat, habitudes de consommation, traits de personnalité ...

⁶Kröger et al., "Surveilling the gamers".

Et le reste ?

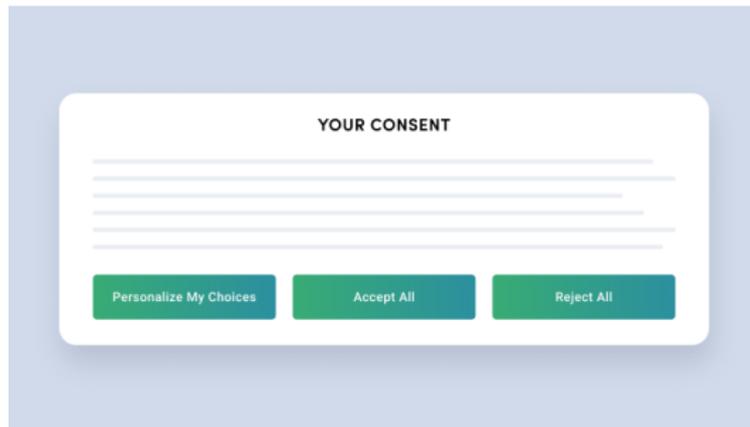
- Quid des autres services, applications, objets connectés ?
 - Vraisemblablement la même chose ...



- 1 Introduction
 - Culture et vie privée
 - Numérisation de l'offre culturelle
 - Collecte et exploitation des données
- 2 "Anonymisation" de profils Netflix
- 3 Applications mobiles de streaming
- 4 Le cas des TV connectées
- 5 Conclusion

Comment se protéger I

- 1 Configurer ses outils et faire respecter ses droits
 - Refuser la collecte de données



Comment se protéger II

- Exercer ses droits : exiger l'effacement de ses données : <https://www.cnil.fr/fr/les-droits-pour-maitriser-vos-donnees-personnelles>

The screenshot shows the CNIL website's page on personal data rights. At the top, there is a navigation bar with links for 'MÉTAMOTRIC', 'GÉNÉRAL', 'RECONCILIATION', 'PRÉVENTION', 'EN', and 'Outils DES USAGERS'. The CNIL logo is prominently displayed, along with a 'PARTAGER' button and a red 'JE SUIS EN PROTECTION' button. Below the logo, a subtitle reads 'Protéger les données personnelles, accompagner l'innovation, protéger les clients individuels'. A secondary navigation bar includes 'MA CONFORMITÉ AU RGPD', 'THÉMATIQUES', 'TECHNOLOGIES', 'TEXTES OFFICIELS', and 'LA CNIL', followed by search and social media icons. The main content area features a home icon and the text 'Comprendre vos droits'. The title of the page is 'Les droits pour maîtriser vos données personnelles'. A paragraph states: 'La CNIL vous informe sur les différents droits que vous pouvez exercer auprès des organismes qui utilisent vos données.' Below this, there are two main sections: 'Rester informé' with a megaphone icon and 'Vous opposer' with a 'OUI NON' sign icon. The 'Rester informé' section explains that organizations collecting information should provide clear information on data use and rights. The 'Vous opposer' section explains that users can oppose their consent if an organization uses their data. A red 'PERSONNES' button is visible in the bottom right corner of the content area.

Comment se protéger III

- 2 Faire des choix
 - Revenir à des solutions "low tech"

Walkman

Sony's Walkmans Give You Carry-Around Sound.

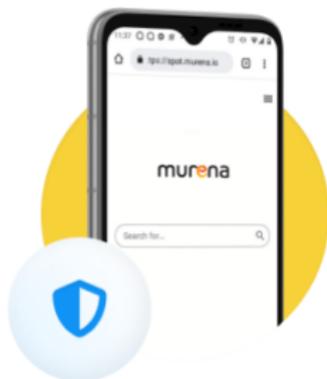
SUPER WALKMAN WM-20

Super Sound with Super Sound. The world's smallest, lightest Walkman, the WM-20 features a built-in stereo speaker and a package no larger than a cassette case. Thanks to Sony's exclusive SBC, makes the WM-20 as just a fifth the size of other Walkmans and 1/10th as heavy. The model delivers crisp, clear audio with dynamic range that's 10 times that of a single CD with battery life that's 10 times longer. The WM-20 is the most powerful Walkman in the world. It's the most powerful Walkman in the world.

Scanned by: Asim Bijarani

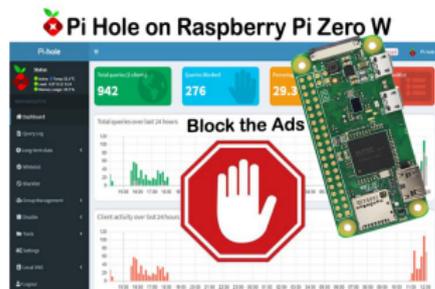
Comment se protéger IV

- Utiliser des solutions alternatives plus vertueuses: Android "degooglisé Murena" : <https://e.foundation/fr/e-os/>



Comment se protéger V

- 5 Utiliser des outils pour la protection et la transparence
 - PrivacyBadger⁷, Pi Hole⁸, Exodus⁹, ...



⁷<https://privacybadger.org/>

⁸<https://pi-hole.net/>

⁹<https://exodus-privacy.eu.org/>

`mathieu.cunche@insa-lyon.fr`
`http://perso.citi-lab.fr/mcunche/index.html`
twitter: @Cunchem

- Kröger, Jacob Leon et al. "Surveilling the gamers: Privacy impacts of the video game industry". en. In: *Entertainment Computing* (Oct. 2022), p. 100537. ISSN: 1875-9521. DOI: 10.1016/j.entcom.2022.100537. (Visited on 11/02/2022).
- Narayanan, Arvind and Vitaly Shmatikov. "Robust De-anonymization of Large Sparse Datasets". In: *2008 IEEE Symposium on Security and Privacy (sp 2008)*. May 2008, pp. 111–125. DOI: 10.1109/SP.2008.33.
- Pasquier, Dominique. "L'espace privé comme lieu de consommation culturelle". fr. In: *Création et diversité au miroir des industries culturelles. Questions de culture*. Paris: Ministère de la Culture - DEPS, 2006, pp. 239–254. ISBN: 978-2-11-096197-6. DOI: 10.3917/deps.gref.2006.01.0237. (Visited on 11/02/2022).
- Ren, Jingjing et al. "Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach". en. In: *Proceedings of the Internet Measurement Conference*. Amsterdam Netherlands: ACM, Oct. 2019, pp. 267–279. ISBN: 978-1-4503-6948-0. DOI: 10.1145/3355369.3355577. (Visited on 11/02/2022).
- Tileria, Marcos and Jorge Blasco. "Watch Over Your TV: A Security and Privacy Analysis of the Android TV Ecosystem". en. In: *Proceedings on Privacy Enhancing Technologies 2022.3* (July 2022), pp. 692–710. ISSN: 2299-0984. DOI: 10.56553/popets-2022-0092. (Visited on 11/02/2022).