

# Internet et Big Brother : Réalité ou Fantôme ?

Dr. Pascal Francq

- Introduction
- Traces
- Google
- Applications
- Enjeux
- Conclusions

- **Introduction**
- Traces
- Google
- Applications
- Enjeux
- Conclusions

- Internet sert de support pour la manipulation de **données à caractère privé** :
  - États (Taxonweb, données médicales, etc).
  - Entreprises (banques, Amazon, etc).
  - Comportements (requêtes Google, navigation, etc).
- Techniques de “data-mining” :
  - Croiser des données provenant de différents systèmes d'information (ex: Banque et société de crédits).
  - Intelligence artificielle permet de **profiler** le comportement des internautes (ex: Recommandations dans Amazon).

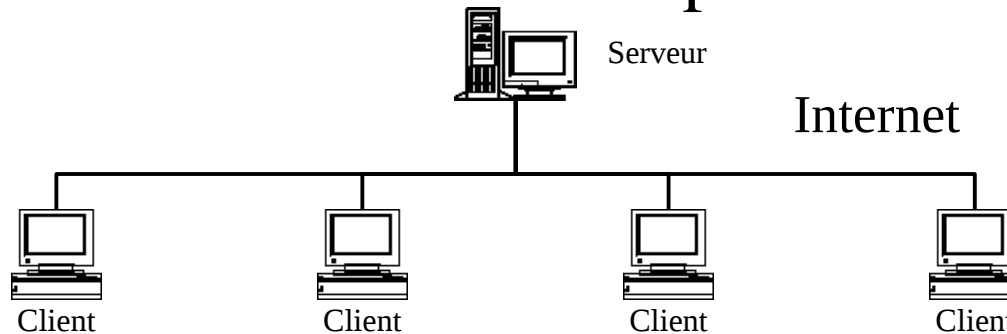
- Introduction
- **Traces**
- Google
- Applications
- Enjeux
- Conclusions

- Chaque machine connectée est identifiée par un numéro unique, **Adresse IP** :
  - Adresse peut être *statique*, c'est-à-dire fixée pour une machine donnée (et donc pour un ou plusieurs utilisateurs).
  - Adresse peut être *dynamique*, c'est-à-dire gérée par un fournisseur d'accès (qui sait toujours à quel client il attribue quelle adresse quand).
  - Café “Internet” garde généralement une information sur ses clients,
- Toute donnée “transportée” par Internet contient les adresses IP de l'émetteur et du récepteur.

## Traces (2/4) - Centralisation d'Internet

---

- Infrastructure basée sur quelques “grands axes” par lesquels passent toutes les données.
- Plupart des fonctionnalités sont basées sur une **architecture client/serveur** :
  - Internauts utilisent des **applications clients** (navigateur, client de messagerie, lecteur de news, etc.) pour faire des requêtes.
  - Services sont hébergés par des serveurs qui **centralisent** les demandes et répondent aux requêtes.



## Traces (3/4) - Fichiers Log

---

- Serveurs peuvent garder les traces de toutes les requêtes qu'ils traitent.
- En pratique, ils gèrent des fichiers log (**log files**) stockant différentes informations :
  - Nom de la machine et son adresse IP.
  - Type, date et heure des requêtes.
  - Certaines informations sur l'utilisateur (nom, e-mail, etc).
  - Certaines informations sur la requête (par exemple les mots-clés utilisés pour une recherche).
- On peut donc tracer le **comportement des internautes**.



- **Cookies** sont des informations “stockées” par un programme (ex: Navigateur) et qui peuvent être lues par un serveur Web.
- Quelques utilisations problématiques :
  - **Suivi du parcours** de l'internaute (quelles pages et quels liens ont été suivis pour arriver sur une page donnée).
  - Dépôt d'informations par une application, puis récupérées par un site Web.
- Cookies peuvent être désactivés dans les navigateurs mais s'est rarement le cas par défaut.

- Introduction
- Traces
- **Google**
- Applications
- Enjeux
- Conclusions

- Google analyse toutes les informations récoltées par ses services :
  - Toutes les requêtes des internautes et les pages recommandées consultées depuis 1998.
  - Contenu des e-mails envoyés via son service de messagerie électronique, Gmail.
  - Contenu des documents générés par ses outils de bureautique en ligne (Google Docs, etc.).
- Google conserve certaines archives de documents mis en ligne.

## Google (2/3) - Exploitation

---

- Officiellement : L'objectif est d'améliorer la qualité des résultats de son moteur.
- En pratique :
  - Personne ne sait exactement ce que Google fait des informations qu'il récolte.
  - Informations sont utilisées par sa régie publicitaire.
- Google a levé 4 milliards de dollars :
  - Volonté de concentrer un maximum d'outils.
  - Bénéfices attendus par l'exploitation des informations collectées.

## Google (3/3) - Données à Caractère Privé

---

- Aucun contrôle des utilisateurs sur ses données manipulées par Google :
  - Aucune possibilité d'effacer ses traces de recherche.
  - Internauts ne savent généralement pas que Google analyse (automatiquement) ses données.
- Aucun contrôle étatique :
  - Commission Européenne a demandé plusieurs fois à Google de respecter les principes en vigueur en Europe en matière de protection des données à caractère personnel.
  - Mais Google est une société américaine dépendant du droit américain et non européen.

- Introduction
- Traces
- Google
- **Applications**
- Enjeux
- Conclusions

## Applications (1/5) - E-mails

---

- Tous les **e-mails échangés** sont stockés sur des serveurs appartenant à des sociétés privées (fournisseurs de services, entreprises de service, etc).
- Exemple : Yahoo a collaboré avec le gouvernement chinois pour identifier et arrêter certains opposants en analysant les messages que ces derniers avaient envoyés.

## Applications (2/5) - Logiciels Sociaux

---

- **Échanger des ressources intéressantes :**
  - Centralisation de qui référence quoi,
  - Les plus connus : Delicious, YouTube.
- **Logiciel de réseautage social :**
  - Centralisation de données sur les personnes (descriptifs, photos, etc) et des liens entre elles.
  - Les plus connus : Facebook, MySpace ou LinkedIn.
- **Exemple : Facebook renseigne des annonceurs en leur fournissant des informations sur ses utilisateurs.**



- Journal de bord en-ligne :
  - Écrit par quelques personnes (parfois une seule).
  - Mises à jour régulières.
  - Liberté de ton et de contenu.
  - Plupart des blogs traitent de la **vie quotidienne** des internautes (affectif, politique, etc.).
- Exemple : Au début de la seconde guerre du Golfe, des journalistes perdirent leur travail car ils défendaient, sur leur blog personnel, des positions non alignées sur la ligne éditoriale de leur journal.

## Applications (4/5) - Agents Intelligents

---

- Certains logiciels sont capables de **construire un curriculum vitae complet** en collectant des données disponibles sur Internet.
- Types d'informations récoltées :
  - Études suivies et les positions professionnelles occupées.
  - Hobbies et opinions politiques.
  - Photos de son baptême folklorique à l'université.
  - etc.
- Exemple : Entreprises s'immiscent dans la vie privée de leurs (futurs) employés.

## Applications (5/5) - Et Demain ?

---

- Multiplication des sites de référencement de personnes (ex: Site "stop kinderporno").
- Créé en 2006, le site Web “spock.com” :
  - Plate-forme permettant de référencer tout humain existant ou ayant existé.
  - N'importe qui peut choisir les mots-clés caractérisant au mieux les personnes reprises sur le site.
- Exemple : Mots-clés associés par les internautes au chanteur français Serge Gainsbourg : “Jewish” et “agnostic”.

- Introduction
- Traces
- Google
- Applications
- **Enjeux**
- Conclusions

- Belgique s'est dotée d'une loi de protection de la vie privée :
  - Commission indépendante.
  - Loi louable mais difficile à mettre à œuvre.
  - Arsenal “répressif” insuffisant.
- Si Internet est mondial, chaque pays a sa propre législation :
  - Pas même niveau de protection.
  - Pas d'uniformisation mondiale.
- Législation mondiale est une utopie,

- Terme “Pirates” regroupe la **cyber-criminalité** ( $\neq$ Hacker).
- Pirates sont généralement de **très bons techniciens**.
- Objectifs :
  - Prendre des informations sur des ordinateurs à des fins de fraudes ou de vols.
  - Détruire des informations.
  - Passer à travers la sécurité pour le plaisir.
- Sensibiliser les entreprises et les internautes à la protection des données informatiques.

### Investissements

- Diminuer la part du privée dans le financement d'Internet.
- Proposer des **services publics numériques** :
  - Mise en place de messageries protégeant le contenu des e-mails.
  - Développement d'un moteur de recherche public.
- Investir dans le développement de nouvelles technologies.

### Éducation

- Former les internautes :
  - Comprendre les dangers liés aux différentes applications.
  - Apprendre à maîtriser les outils pour mieux se protéger (ex: Fonctions avancées des navigateurs).
- Cible privilégiée doit être les jeunes : Introduction d'un cours “**d'éducation numérique**” dès le secondaire,



- Introduction
- Traces
- Google
- Applications
- Enjeux
- **Conclusions**

- Outils d'Internet permettent de récolter toutes sortes d'informations sur les internautes :
  - Qui lit quoi ?
  - Qui fait quelles recherches ?
  - Qui correspond avec qui ?
  - Qui publie quoi ?
- Principaux outils sont aux mains de quelques entreprises privées dont les recettes dépendent de leur “connaissance” des internautes.
- Google ou Facebook illustrent bien les enjeux de la protection des données à caractère privé.