Economie de la vie privée

Journée CAPPRIS - Lyon - 10 septembre 2013

Fabrice Le Guel, Matthieu Manant

fabrice.le-guel@u-psud.fr , matthieu.manant@u-psud.fr

ADIS - Université Paris Sud

ANR ESPRI

http://www.adis.u-psud.fr/espri/

UNIVERSITÉ PARIS SUD

FACULTÉ JEAN MONNET
Droit - Économie - Gestion

ADIS UNIVERSITE PARIS SUD-11

Agence Nationale de la Recherche ANR

Le laboratoire ADIS (PESOR)

- Laboratoire en économie de l'Université Paris Sud
- Une équipe 'économie des TIC' (6 chercheurs permanents)
- Un Master 'Ingénierie de l'Innovation, Valorisation de la Recherche et Transferts de Compétences' (IVRTC)
- Un Master 'Industries de Réseau et Économie Numérique' (IREN)

Notre savoir-faire

- **Traitement de l'information (.Net, Python, R, Stata, Matlab, SAS)**
- **Base de données (XML, ACCESS, MySQL)**
- **Web scraping**
- **Analyse de données**
- **Statistiques**
- **Econométrie**
- **Analyse des interrelations (graphes relationnels)**
- **Economie expérimentale (une salle disponible en permanence)**
- **Field experiment**
- **Enquêtes, sondages (ex: comportement de piratage des français)**
- **Economie computationnelle - Projet LACOSS (Laboratoire Computationnel en Sciences Sociales) – programmation multi-agents.**
- **Economie de la vie privée, tracking (Internet, iPhone, Android…)**

Privacy/Vie privée : définitions

- Warren et Brandeis, 1890 :

"the right to be let alone"

- Westin, 1967 :

"Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."

- Stone et al., 1983 :

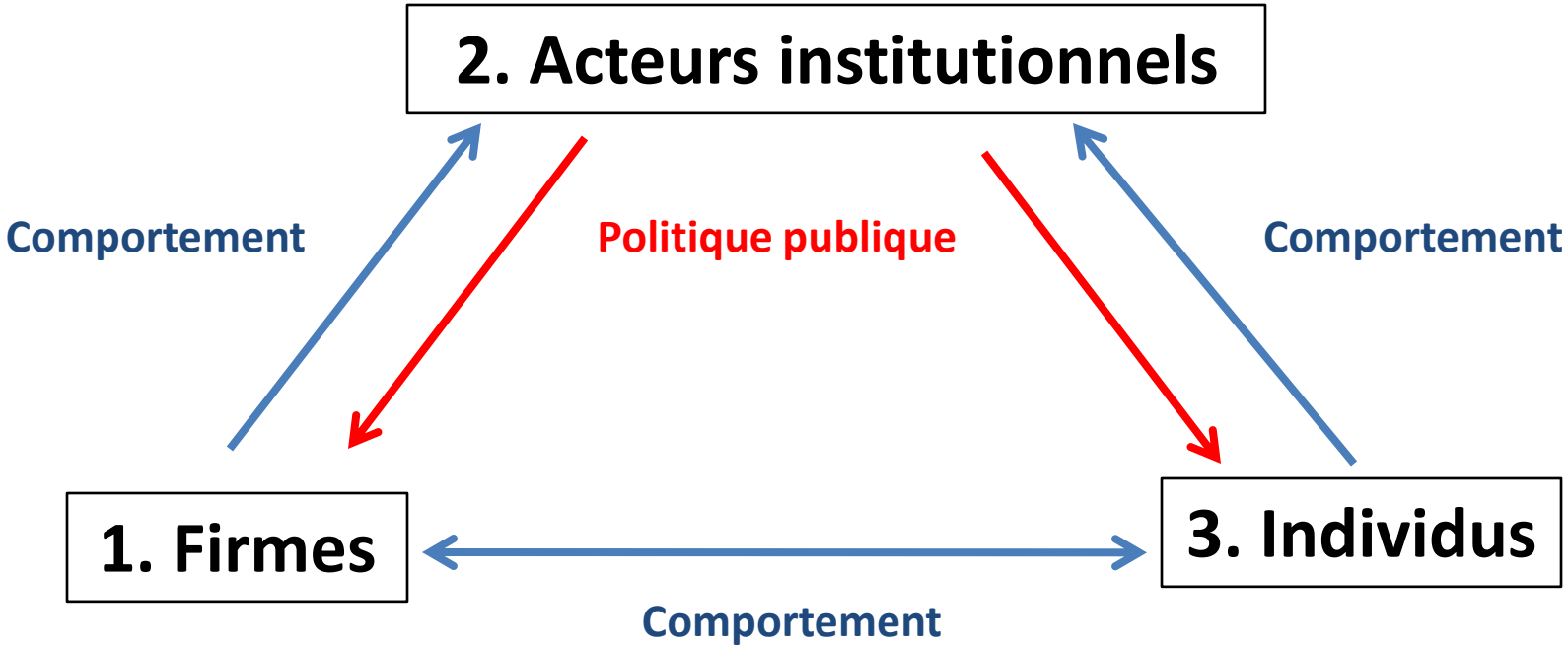
« […] the ability of the individual to personally control information about one-self »

- In Hui and Png, 2006 :

"Information privacy has been defined as the individual's ability to control the collection and use of personal information (Westin, 1967; Stigler, 1980)."

L'économie de la vie privée

Nous analysons 3 types d'acteurs :



Privacy/Vie privée : la régulation

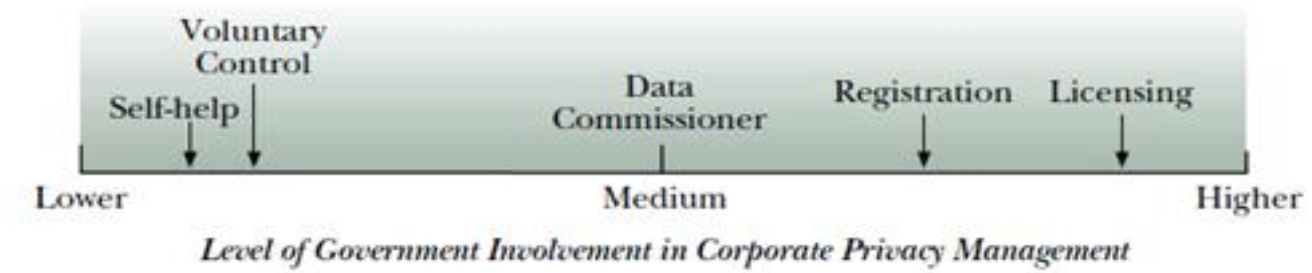


Figure 1. Regulation Models. The models can be described as follows:

1)The "Self-Help" Model depends on data subjects' challenging inappropriate record-keeping practices. Rights of access and correction are provided for the subjects, but they are responsible for identifying problems and bringing them to the courts for resolution.

2) The Voluntary Control Model relies on self-regulation on the part of corporate players. The law defines specific rules and requires that a "responsible person" in each organization ensures compliance.

3) The Data Commissioner Model utilizes neither licensing nor registration, but relies on the ombudsman concept through a commissioner's office. The commissioner has no powers of regulation but relies on complaints from citizens, which are investigated. The commissioner also is viewed as an expert who should offer advice on data handling; monitor technology and make proposals; and perform some inspections of data processing operations. This model relies to a great degree on the commissioner's credibility with the legislature, the press, and the public.

4) The Registration Model acts much like the Licensing Model, with one exception: the government institution has no right to block the creation of a particular information system. Only in a case where complaints are received and an investigation reveals a failure to adhere to data protection principles would a system be "deregistered." Thus, this model provides more remedial than anticipatory enforcement of principles.

5) The Licensing Model creates a requirement that each databank containing personal data be licensed (usually upon payment of a fee) by a separate government institution. This institution would stipulate specific conditions for the collection, storage, and use of personal data. This model anticipates potential problems and heads them off, by requiring prior approval for any use of data.

*Adapted from Smith [17]; categories based on Bennett [2].

Source : Milberg et al (1995)

Privacy/Vie privée : la régulation

Table 2 Country Classifications

Classification	Countries
No *formal* information privacy regulation[a]	Argentina Italy India Malaysia Thailand
Self-Help (Model 1)	None
Voluntary Control (Model 2)	Japan United States
Data Commissioner (Model 3)	Australia Canada Finland Germany New Zealand Switzerland
Registration (Model 4)	Belgium Denmark France Netherlands U.K./England/Scotland
Licensing (Model 5)	Norway

Source : Milberg et al (2000)

Le consommateur

Individual privacy concerns

Smith et al. (1996) ont identifié 4 dimensions qui permettraient de tester le niveau de 'privacy concerns' des individus :

- [*Collection*] Abus dans le volume collecté de données,
- [*Unauthorized internally/externally secondary use*] Usage secondaire des données personnelles (avec un usage secondaire 'interne', i.e. au sein de l'entreprise et un usage secondaire 'externe' à l'entreprise, i.e. les 'third parties'),
- [*Improper access*] Accès non authorisé aux données personnelles au sein d'une entreprise (i.e. un employé accède à des données personnelles sans autorisation : banques, police…),
- [*Errors*] Accidents dans la saisie des données personnelles avec des difficultés pour corriger l'erreur.

Le consommateur

Concern for information privacy scale items

Privacy scale factors and individual scale items
(1 strongly disagree–7 strongly agree)

Collection
1. It usually bothers me when Web sites ask me for personal information.
2. When Web sites ask me for personal information, I sometimes think twice before providing it.
3. It bothers me to give personal information to so many Web sites.
4. I'm concerned that Web sites are collecting too much personal information about me.

Improper access
1. Web sites should devote more time and effort to preventing illegal access to personal information.
2. Databases that contain personal information should be protected from illegal access—no matter how much it costs.
3. Web sites and other companies should take more steps to make sure that hackers cannot access the personal information in their computers.[a]

Errors
1. All the information received on Web sites should be double-checked for accuracy—no matter how much this costs.
2. Web sites should take more steps to make sure that the personal information in their files is accurate.
3. Web sites should have better procedures to correct errors in personal information.
4. Web sites should devote more time and effort to verifying the accuracy of the personal information in their databases.

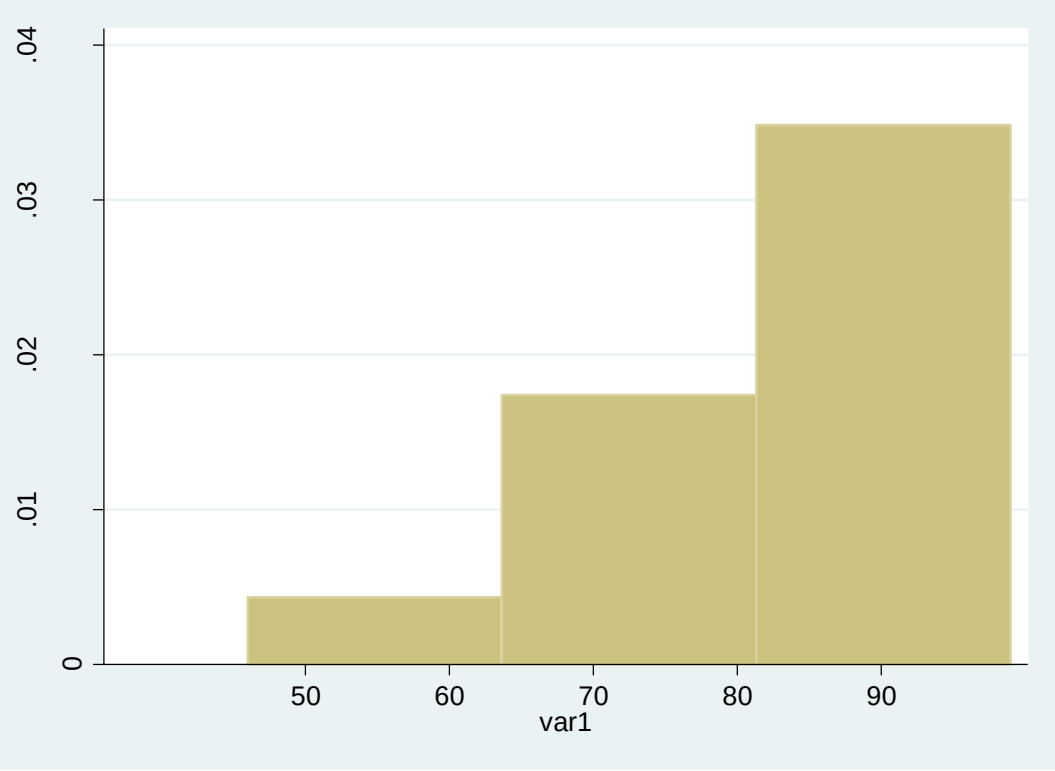
Unauthorized secondary use
1. Web sites should not use personal information for any purpose unless it has been authorized by the individuals who provide the information.
2. When people give personal information to a Web site for some reason, the Web site should never use the information for any other reason.
3. Web sites should never sell the personal information they have collected to other Web sites.
4. Web sites should never share personal information with other Web sites or companies unless it has been authorized by the individuals who provided the information.

Smith et al. (1996) ont alors proposé un questionnaire standard pour mesurer le niveau de 'privacy concerns' de chacun, que Bellman et al. 2004 ont ici adapté aux sites Internet:

Source : Bellman et al (2004)

Le consommateur et le ‘privacy paradox’

Etes-vous ‘privacy concerns’ ? (100 % des étudiants ont un compte Facebook)

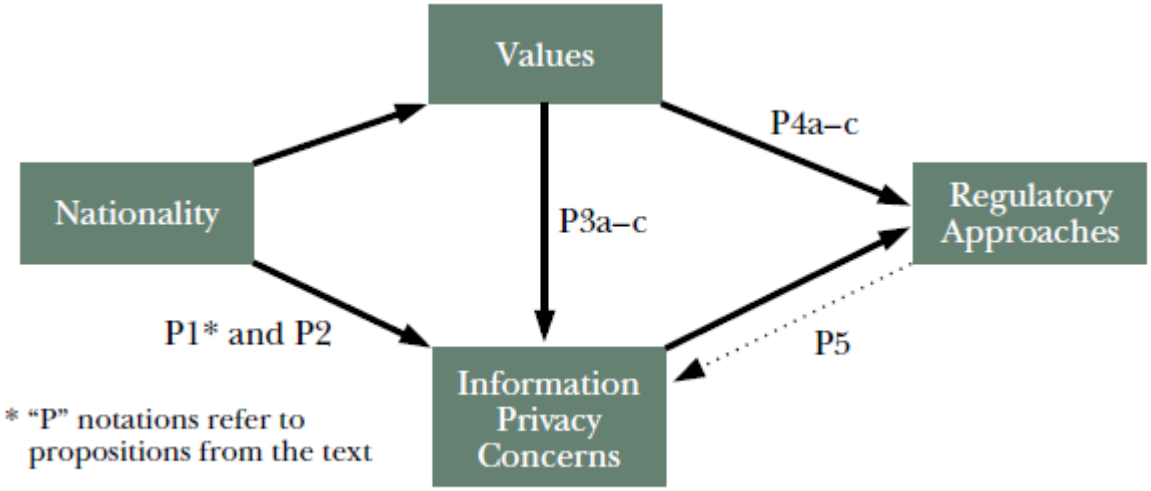


Variable	Obs	Mean	Std. Dev.	Min	Max
var1	45	81.30769	14.29676	46	99

Mais le consommateur est le premier fournisseur d'information !

Remettre le consommateur au centre

Des interactions constantes

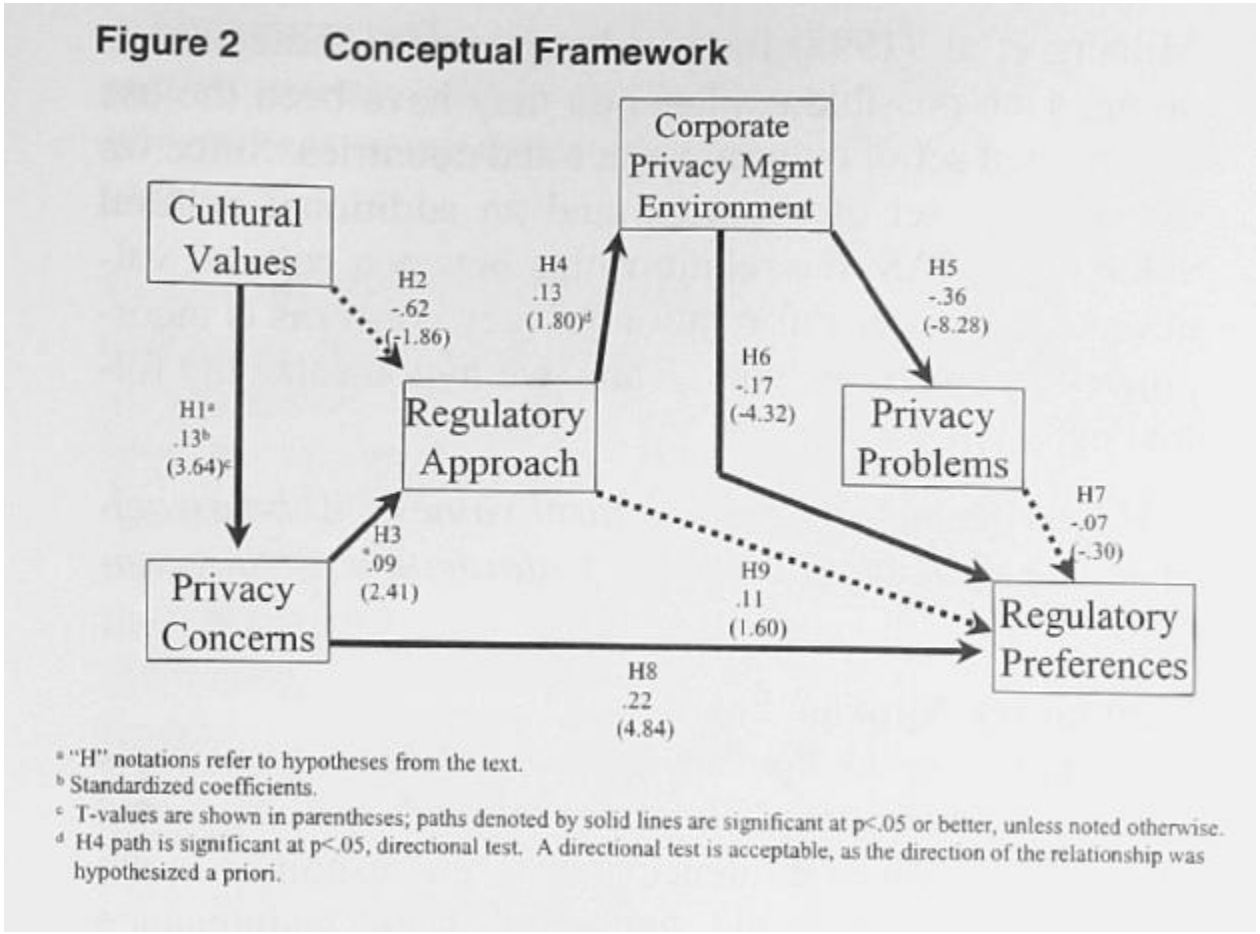


Source Milberg et al. 1995

Des variations temporelles et spatiales !

Privacy/Vie privée : définitions

Des interactions constantes:



Source Milberg et al. 2000

L'économie de la vie privée

Analyse économique de 3 types d'acteurs en prenant davantage en compte les consommateurs :

- Acteurs institutionnels

Analyse de l'action des acteurs institutionnels et des politiques publiques

- Firmes

Analyse du comportement des firmes, métrique.

- **Individus**

Analyse de la prise de décision des individus

Privacy/Vie privée : définitions

Un arbitrage toujours délicat entre :

Respect de la vie privée ⟺ Surveillance ⟺ Divulgation (disclosure)

Question clef : arbitrage entre libre circulation des données personnelles et protection des données personnelles (faut-il réguler et comment ?)

ANR ‘ESPRI’ (3 ans – Fin Novembre 2013)

http://www.adis.u-psud.fr/espri/

- ESPRI ("Exposition de soi, Privacy et Réseaux d'interaction") est un projet de recherche pluridisciplinaire économie-sociologie-droit financé par l'ANR.

- Objectif : clairer l’ mergence et les s de alisation

  la vie e en se fondant sur une analyse des comportements et des pratiques elles des offreurs de services et des utilisateurs de ces services.

ANR ESPRI

- **Vie privée et labels (field experiment)**

- **Discrimination à l'embauche et vie privée (field experiment)**

- **La mesure de la vie privée et les différences culturelles (base de données européenne)**

- **Vie privée et localisation des serveurs Internet (scraping de données)**

- **Capacité à révéler de l'information personnelle (expérience Quizz)**

- **Etude du marché des données personnelles sur Smartphone et du comportement des firmes (tracking de données)**

- **Comportement des utilisateurs de smartphone face à une intrusion dans la vie privée (field experiment)**

Le projet ‘Privacy Apps’

Composé de deux parties :

- **‘Privacy Apps phase I’: (le comportement des firmes)**

Analyse du comportement des firmes (3 acteurs : ceux qui possèdent l’OS Apple *versus* Android, les applications, les acteurs tiers) sur les smartphones , en matière d’utilisation des données personnelles

- **‘Privacy Apps phase II’: (le comportement des utilisateurs)**

Analyse du comportement des utilisateurs de smartphone lorsqu’ils sont confrontés aux questions de vie privée

Phase I : Measuring the ‘Invisible Economy’ of Smartphone Applications

Context :

Scott Thurm and Yukari Iwatani Kane's survey

'Your Apps Are Watching You'

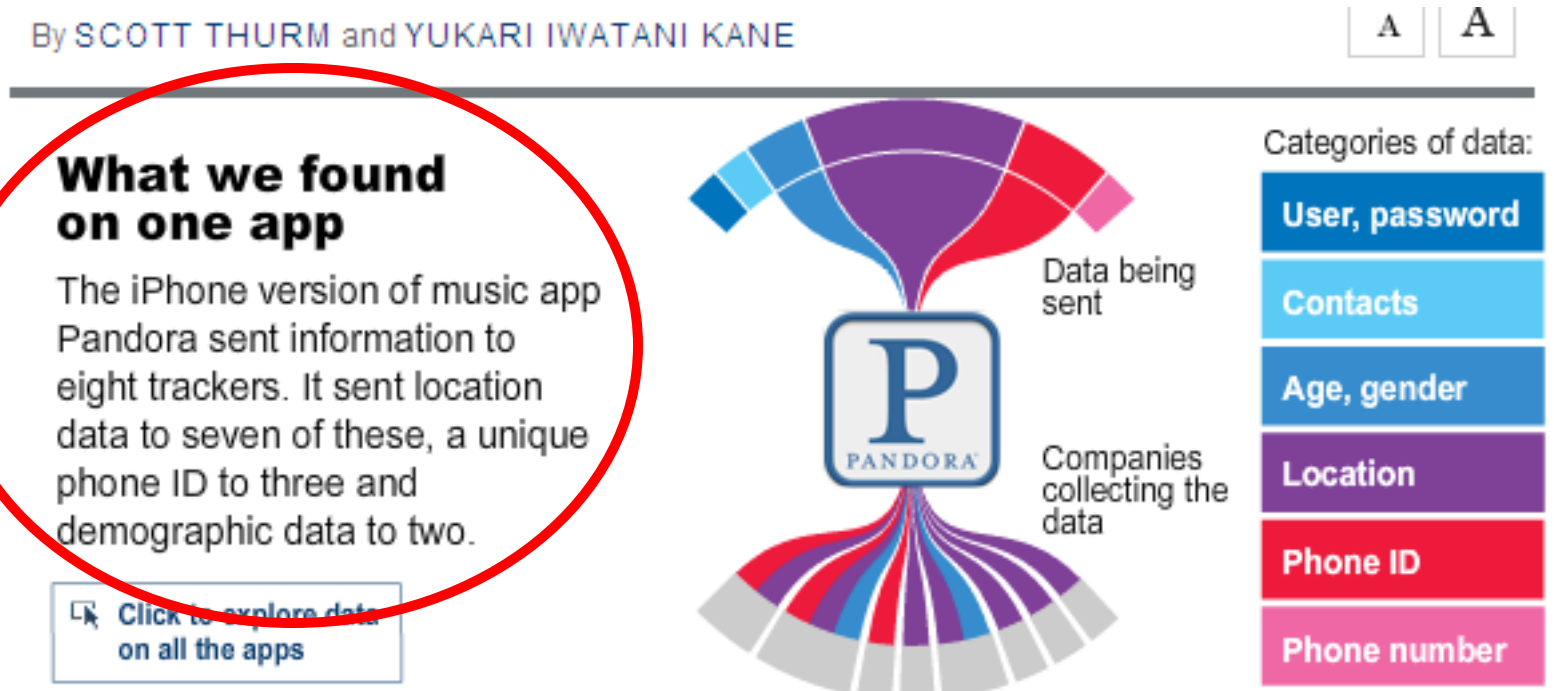
The Wall Street Journal, December 17, 2010

- **WSJ's authors observed the transmission of personal information of 101 popular US smartphone applications on iPhone and Android devices**

http://blogs.wsj.com/wtk-mobile/

Context :

- **Authors found that iPhone and Android apps were breaching the privacy of smartphone users**



http://blogs.wsj.com/wtk-mobile/

Context :

- **So Thurm and Kane showed that personal information is transmitted by most applications *without the user being aware of it !***

- ***Particularly, an application can transmit personal information:***

1. ***to the application owner/developer***
2. ***to Apple or Google, through iphone or Android devices : the device owner***
3. ***to third parties (i.e. tracking companies, marketers,* data aggregators *…)***

Literature
(in privacy economics) :

- **App owner/device owner** (when gathering personal information), corresponds to a « **first usage** of personal information » (Varian, 1996), also called « **primary market** of personal data » (Akçura & Srinivasan, 2005)

- **Third party** (when gathering personal information from app owner(s)) corresponds to a « **second usage** of personal information » (Varian, 1996), also called « **secondary market** of personal data » (Akçura & Srinivasan, 2005) or **cross-selling market**.

Problem

Akçura & Srinivasan, (2005) :

- **'first market'/'first usage' may help app owner to better interact with customers, to better target offers, etc…**

- **…BUT a 'secondary market'/'second usage'/ 'cross-selling market' may dramatically decresase consumers' welfare (spamming, monitoring, profiling…)**

The aims of the paper:

- **is to forecast the size of this secondary market (here named the 'invisible economy' - *from the point of view of the consumer*)**

- **and evaluate the market concentration (i.e. the distribution of market shares)**

- **Policy regulation**

<u>With this assumption :</u>

Assumption :

The ‘visible economy’ of smartphone apps : personal data are transmitted with the user’s agreement

The ‘invisible economy’ of smartphone apps : personal data are transmitted without consumers being aware of it

I suppose that an “invisible economy” of smartphone applications is growing and is becoming at least as important as the visible one

Data :

- I use data from Scott Thurm and Yukari Iwatani Kane's survey (TheWall Street Journal, December 2010)

- The sample contains exactly 50 applications on Apple's App Store, 50 applications on the Google's Android Market, plus the iPhone application of the Wall Street Journal

- Each application had been used during about 5 minutes, and personal data sent had been tracked.

Table 1. Amount of information sent per application

Personal information	Number of observations	Mean	Standard deviation	Min.	Max
User name, password	101	0.39	0.62	0	3
Contacts	101	0.04	0.20	0	1
Age, gender	101	0.14	0.57	0	3
Location	101	1.21	1.57	0	7
Phone id	101	2.04	2.31	0	10
Phone number	101	0.08	0.31	0	2

Methodology :

Unlike the Thurm and Kane's paper:

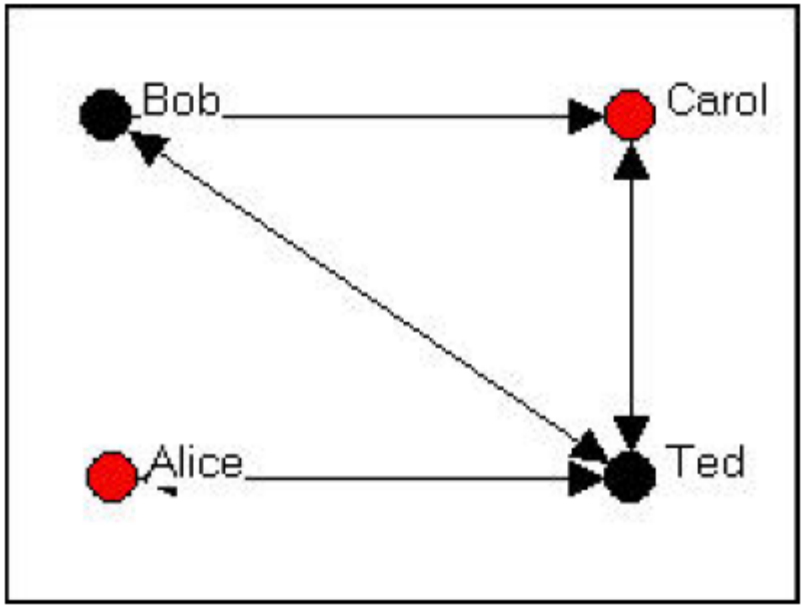
- **I use relational graphs and statistics to forecast the amount of competition between companies collecting personal data (market concentration),**

- **and I use econometrics to test the determinants of the number of personal data sent**

First result

It is possible to better analyze transmission of personal data with relational graphs

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---



First result

It is possible to better analyze transmission of personal data with relational graphs

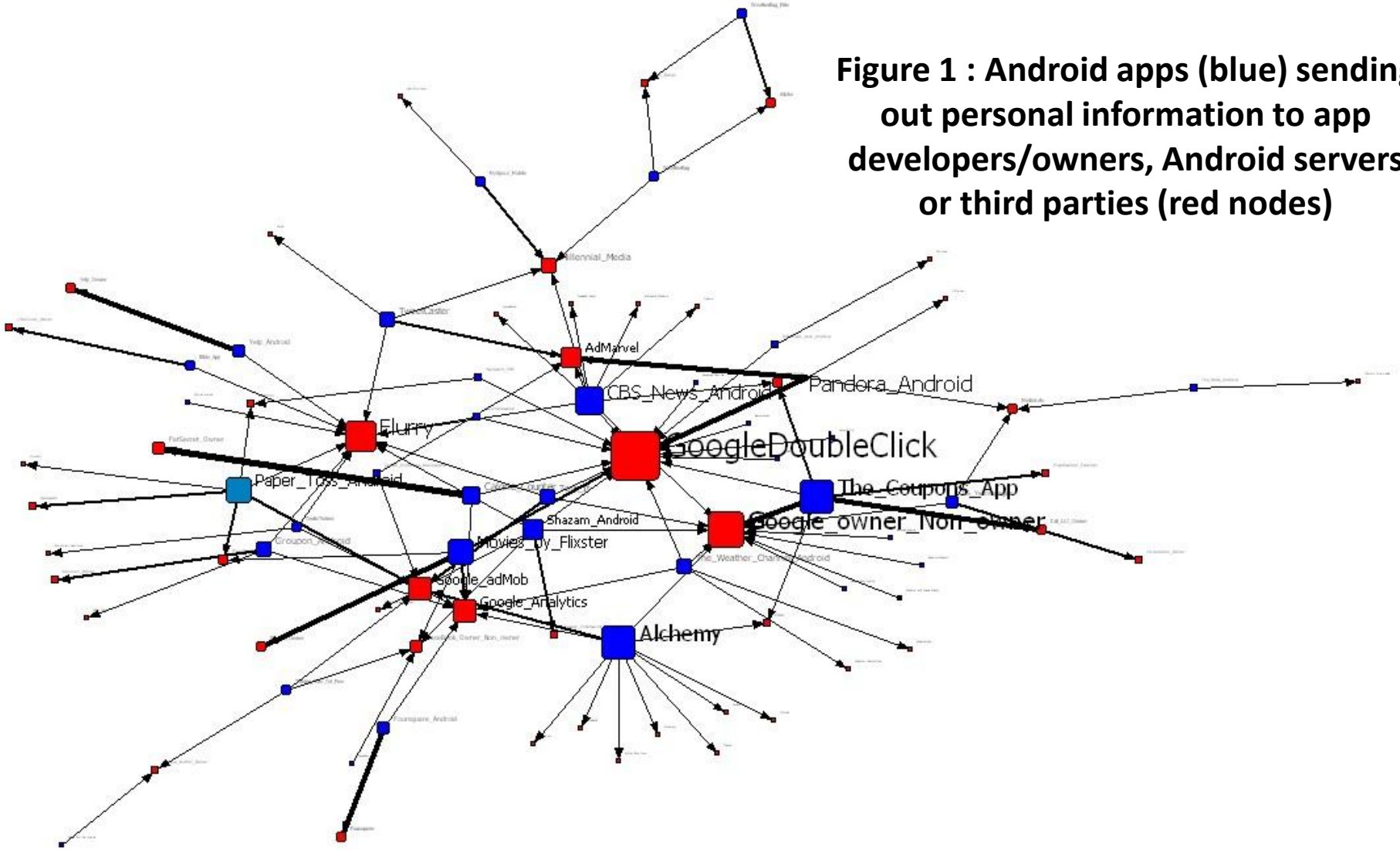
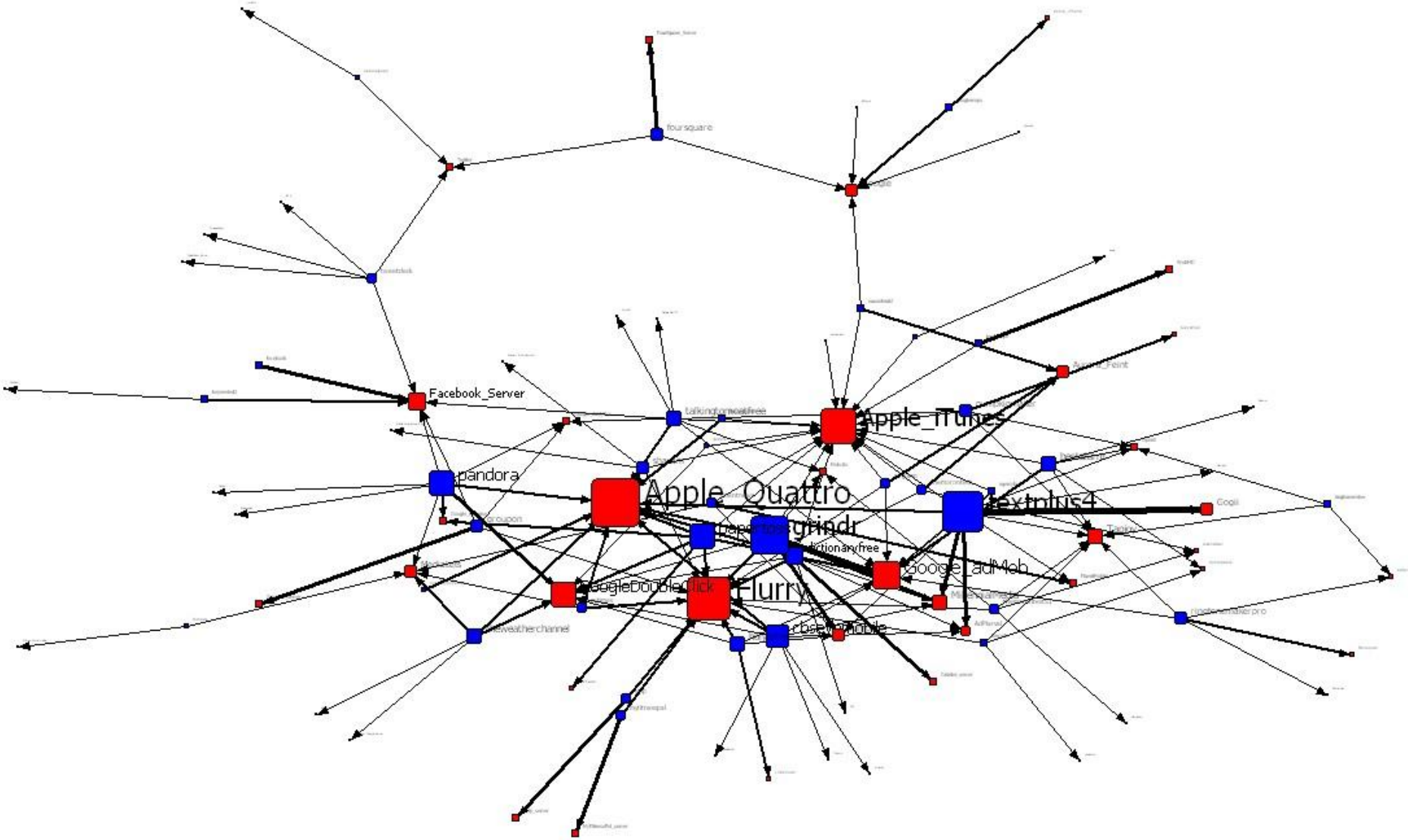


Figure 1 : Android apps (blue) sending out personal information to app developers/owners, Android servers or third parties (red nodes)

First result

Figure 2 : iPhone apps sending out personal information to app developers/owners, Apple servers or third parties



First result

On the Android platform:

- **44 % of the applications sent personal information to the application developer (primary market) (55% for the iPhone)**

- **and 77 % of the applications transmitted personal data to the so-called 'third parties' (secondary market), i.e. to data aggregators (78% on the iPhone).**

Second result

The market structure of personal data seems to be hightly concentrated (according to a Zipf distribution and the Herfindahl index)

For example : the 50 Android applications transmitted 150 personal data to 53 servers, but only the first 8 applications (in terms of quantity of data transmitted) sent about 50 % of the whole 150 data transmitted

Figure 9 : Rank-size distribution of received data by servers on Android OS

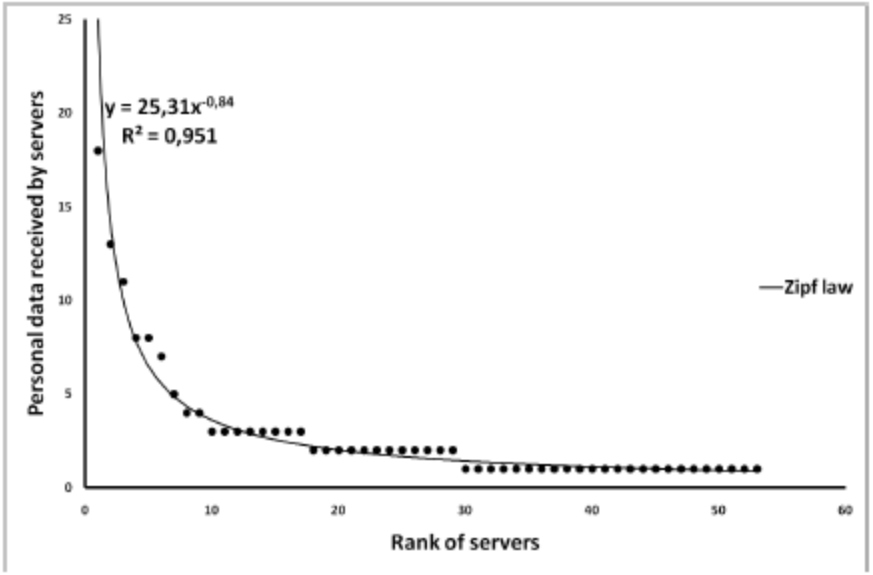
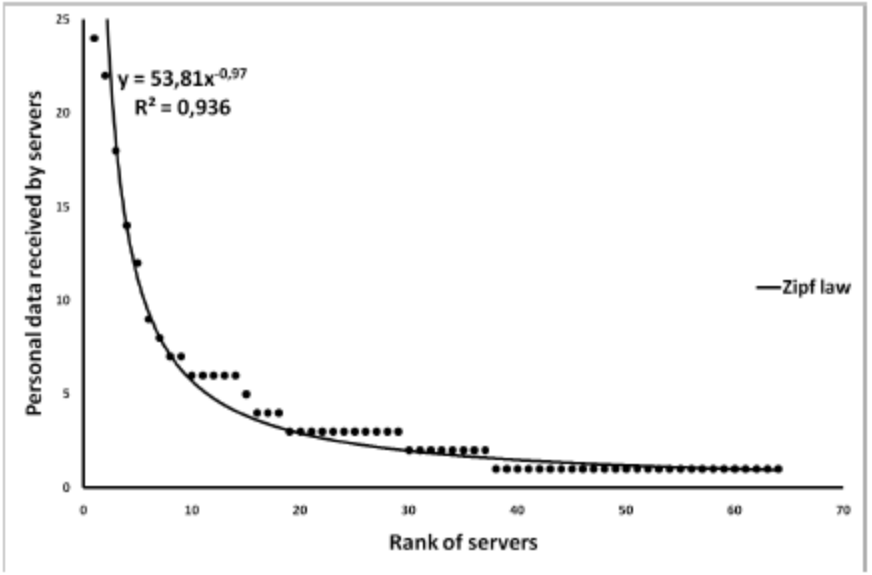


Figure 10 : Rank-size distribution of received data by servers on iPhone OS



Third result

Count data econometric models to study determinants of the number of personal data sent show that :

- a paid application (versus a free one) is associated with a 50% decrease in the mean number of personal information items sent,

- when an application has a privacy policy, this raises the mean number of personal data sent by 42%,

- If an application is installed on the iPhone (versus the Android emulator), this increases the mean amount of personal information transmitted by 57%,

- when applications transmit personal information to third parties, this increases the number of personal data by 502%

Recommandations

It may be important to:

- better observe the ‘behavior’ of smartphone applications to measure market concentration and thus to impose regulations for higher or weaker market concentration

- better inform consumers about the relationships between smartphone apps and third parties,

- encourage the development of Privacy Enhancing Technologies (PET)

Further developments

- **Next step will aim to confirm results with a larger sample size (first results with a new sample of more than 700 applications confirm our results)**

- **and to study consumers' behavior when faced with privacy concerns on Smartphone**

See the ANR ESPRI project (Self exhibition, Privacy and Social Networks): www.adis.u-psud.fr/espri

Many thanks for your attention !

http://www.keynote.com/benchmark/SaaS/article_web_privacy.shtml

‘Privacy Apps phase II’

L’enquête :

Analyse du comportement des individus lorsqu’ils sont confrontés à la diffusion de données privatives sur smartphone

Questions

- **Les utilisateurs ont-ils conscience du phénomène ?**

- **S'ils ne sont pas conscients du phénomène, diminuer cette asymétrie d'information va-t-il conduire à une modification des comportements ?**

- **S'ils sont conscients du phénomène, est-ce qu'ils perçoivent l'échange de données comme un risque ?**

Hypothèse

Les utilisateurs d'application ont une sensibilité différente face à la diffusion des données personnelles

<u>Corolaire :</u>

- **certains utilisateurs vont s'auto-réguler (modification des comportements),**
- **d'autres voudront une 'protection' légale,**
- **d'autres pourraient exiger une compensation monétaire**

Principe :

Une enquête en panel: étudiants de SUPELEC, Telecom Paris Tech et Université Paris Sud

Principe :

- **Développement d'une application sous Android (un service) permettant d'importer :**

1. **Les fichiers logs du téléphone**
2. **La géolocalisation**
3. **Les usages de connexion/déconnexion du GPS et Data**
4. **Les applications utilisées sur le smartphone**

Protocole expérimental

Questionnaire de début d'expérience + Installation de l'application sur les Android

Observation des usages du téléphone pendant 2 mois

GPS
Wifi
Data
Applications utilisées

Choc informationnel : le document de la CNIL et/ou niveau d'intrusion des applications

Observation des usages du téléphone pendant 2 mois

GPS
Wifi
Data
Applications utilisées

Questionnaire de fin d'expérience

LE CHOC INFORMATIONNEL 1 : les recommandations de la CNIL

• N'enregistrez pas d'informations confidentielles (codes secrets, codes d'accès, coordonnées bancaires…) dans votre smartphone (vol, piratage, usurpation d'identité…).

• Ne désactivez pas le code PIN et changez celui proposé par défaut. Choisissez un code compliqué. Pas votre date de naissance !

• Mettez en place un délai de verrouillage automatique du téléphone. En plus du code PIN, il permet de rendre inactif (verrouiller) le téléphone au bout d'un certain temps. Cela empêche la consultation des informations contenues dans le téléphone en cas de perte ou de vol.

• Activez si possible le chiffrement des sauvegardes du téléphone. Pour cela, utilisez les réglages de la plate-forme avec laquelle vous connectez le téléphone. Cette manipulation garantira que personne ne sera en mesure d'utiliser vos données sans le mot de passe que vous avez défini.

• Installez un antivirus quand cela est possible.

• Notez le numéro "IMEI" du téléphone pour le bloquer en cas de perte ou de vol.

• Ne téléchargez pas d'application de sources inconnues. Privilégiez les plateformes officielles.

• Vérifiez à quelles données contenues dans votre smartphone l'application que vous installez va avoir accès.

• Lisez les conditions d'utilisation d'un service avant de l'installer. Les avis des autres utilisateurs peuvent également être utiles !

• Réglez les paramètres au sein du téléphone ou dans les applications de géolocalisation (Twitter, Foursquare, Plyce...) afin de toujours contrôler quand et par qui vous voulez être géolocalisé. Désactivez le GPS ou le Wi-Fi quand vous ne vous servez plus d'une application de géolocalisation.

CHOC INFORMATIONNEL 2 : les recommandations de la CNIL

L'application AVAST sur Android