”On-Line Privacy Concerns - The Data Retention Directive in a Larger Perspective”

Dag Johansen
Dept. of Computer Science
University of Tromsø

Bergen, 14th May 2008

Privacy is Already at Risk

- Considerable amounts of personal data are already collected and used, with and without user consent.
- We have no idea how much of our personal life already is readily available to government, commercial, and illegal activities.

Outline

(A) Why is personal data interesting?

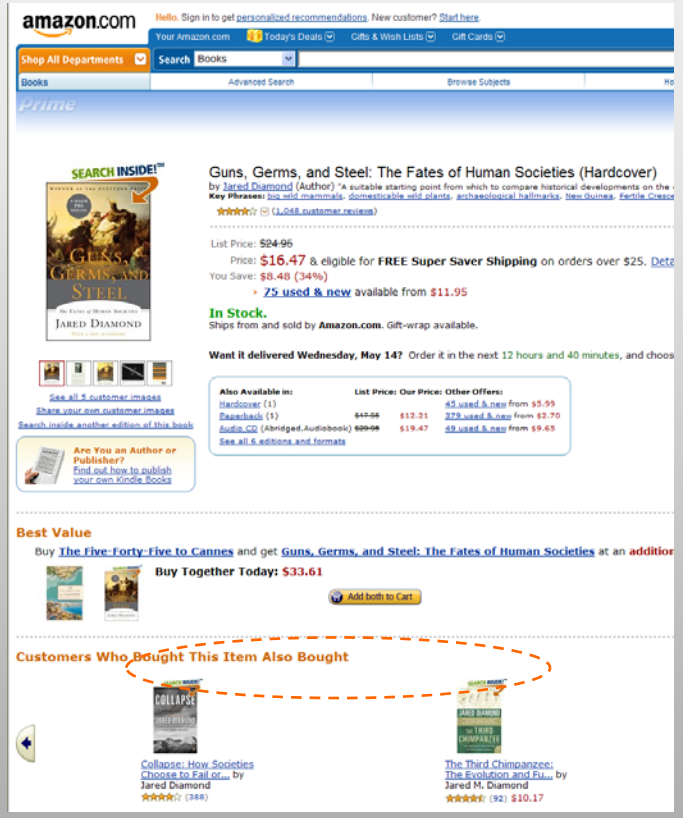
(B) What type of personal data is captured?

(C) The promising technology?

(D) Technical obstacles.

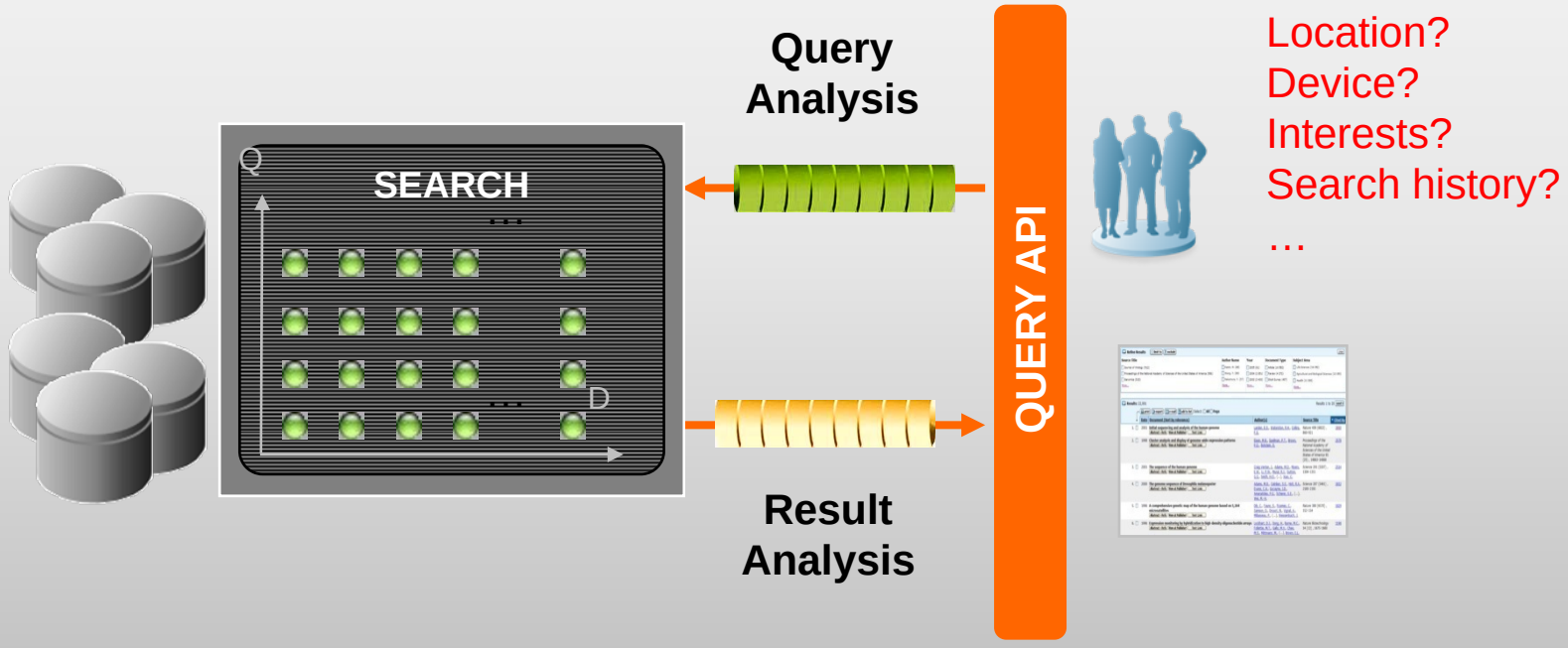
(A) Why is Personal Data Interesting?

- Technical: improved quality and precision in dialogue between users and Internet services:
  - Publish/subscribe and upstream evaluation (RSS feeds, …..).
  - Recommendations.



(1) Technical Reasons

- Precision in search: personal and context sensitive.



(2) Economic Incentive

- On-line advertisement fuels digital economy.
  - IDC: digital economy roughly 30 Billion USD (2006), search based revenue is 50%.
- Advertisement precision; personal data is used to display better ads which in turn means they can earn more money when users click on the ads.
  - ComScore: “Data transmission events” — times when consumer data was zapped back to the Web companies’ servers. Five large Web operations — Yahoo, Google, Microsoft, AOL and MySpace — record at least 336 billion transmission events in a month, not counting their ad networks”.

NYT, March 10th 2008

(3) Illegal Activities

- Improved spam precision.
- Security fraud:
  - Identity theft.
  - Blackmailing.
  - Industrial espionage.
  - …….

(4) Law Enforcement and Control

- Government surveillance:
  - Communication monitoring.
  - Transactions that move money across the border.
  - Use of (Norwegian) credit cards abroad.
  - Digital surveillance cameras.
  - Entry point technology being developed:
    - On-line face recognition, behavioral profiling, and intent capturing (Project Hostile Intent, U.S. Department of Homeland Security).
  - ……
- The European Union data retention directive is yet another mechanism for capturing of personal communications data for law enforcement and control.

(2) What Type of Personal Data?

- 45 Gigabyte/user on average in digital universe; ~50% user generated.

"The Expanding Digital Universe.", IDC 2007.

- "Digital shadow" created by other entities:
  - Program logs.
  - Mailing lists.
  - Digital surveillance cameras.
  - Implicit profiles.
  - Credit records.
  - Web surfing history.
  - …..

→ Anarchy, little control of your personal data despite legal regulations (in many countries).

Examples

”Google collects *personal information* when you register for a Google service or otherwise voluntarily provide such information. We may combine personal information collected from you with information from other Google services or third parties to provide a better user experience, including customizing content for you.

… We may *share aggregated non-personal information* with third parties outside of Google.”

”MySpace.com also collects other profile data including but not limited to: *personal interests, gender, age, education and occupation.*

… MySpace.com also logs non-personally-identifiable information including IP address, profile information, aggregate user data, and browser type, from users and visitors to the site This non-personally-identifiable information may be *shared with third-parties* to provide more relevant services and advertisements to members.”

”It can’t be that harmful ...”

- 20 million Web search queries collected by AOL and released on the Internet to benefit academic researchers.
- New York Times, August 2006: traced AOL id: #4417749.

“numb fingers”

"dog that urinates on everything"

“landscapers in Lilburn, Ga”

“homes sold in shadow lake gwinnett county, georgia”

“60 single men”

” Arnold”*

→ Thelma Arnold, 62, widow, Lilburn, Georgia:

“Those are my searches, ……My goodness, it’s my whole personal life.”

Personal Data Paradox

- EU are investigating commercial companies for storing and using personal data.
- EU themselves are now forcing other commercial companies to store personal data so that they (whatever that means) can use it to monitor private users.

(3) The Promising Technology?

- Billing data are already vital for providing evidence of associations between potential individual criminals and can place them in a particular location.
- Based on cooperation and more tedious insight procedures, now replaced by:
  - Uniform collection procedures.
  - Global snapshot (less missing data).
  - Easier and default access procedures.
  - International collaboration.
  - Improved technology.

”Connection Graf”

- Conjecture: No silver bullet for apriori illegal activity detection based on pure meta-data.

Implications

- Content data (communication) must still be captured through additional targeted surveillance.
- Contact with somebody involved in potentially suspicios activity: ”proof of innocence principle” applies?
- What about effect on confidential political, professional and business communications and contacts?

The Thin Red Line

- Who will have access to the stored information?
  - Local government authorities or non-criminal enforcement authorities who mine into this data and use it for their own governmental (non criminal or counter-terrorism) purposes?
- “They can, why can’t we?”

  “Google said it would anonymise personal data it receives from users' web search after 18 to 24 months. At the time, the firm said it was taking the step partly to *match data retention laws* being rolled out across Europe”.

  BBC News, 25. May 2007

(4) Technical Obstacles

- Can circumvent the system:
  - IP telephony.
  - Skype.
  - Satellite phones.
  - Internet chat.
  - Public phones.
  - Internet cafes.
  - Open wireless networks.
  - Onion routing (anonymous communication networks).

(4) Technical Obstacles

- NAT (several users sharing same IP address).
- Scale:
  - Billions of users.
  - E-mail spam volume.
  - …..

Cost (IKT-Norway: 250 mill. Nkr,1% of revenue), small vendors problem.

- False positives; how to determine that something is wrong:
  - Number of connections?
  - Connections abroad?
  - Connections to specific suspicious people/countries?

(4) Technical Obstacles

- Security; I do not trust private companies with:
  - Many employees.
  - Backup routines (backup copy lost).
  - Weak access control and logging mechanisms.
  - Encryption regimes.
  - …..
- ”Talkmore”:
  - 150.000 Nkr fine when 100.000+ customer records were stolen last year (and they did not notify the authorities).

Concluding Remarks

- How do we balance law enforcement's and commercial needs with (what's left of) privacy when technology permeates all kind of personal data collection?
- Technical constraints and problems, so what will be the next step in mass surveillance?
- Primarily a principal concern:
  - Yet a privacy invasion that can be misused.
  - Principle of ”innocence until proven guilty”.

Questions?