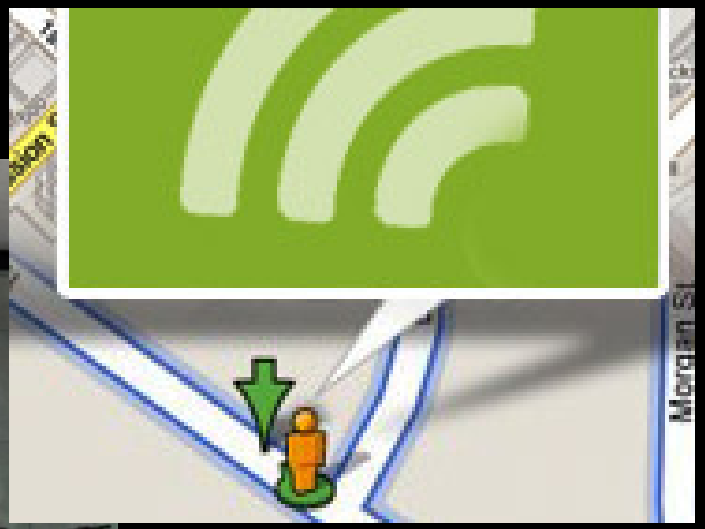
MASSIVE DATA COLLECTION BY MISTAKE?

Arnold Roosendaal
TILT, Tilburg University

Introduction

What is going on here?

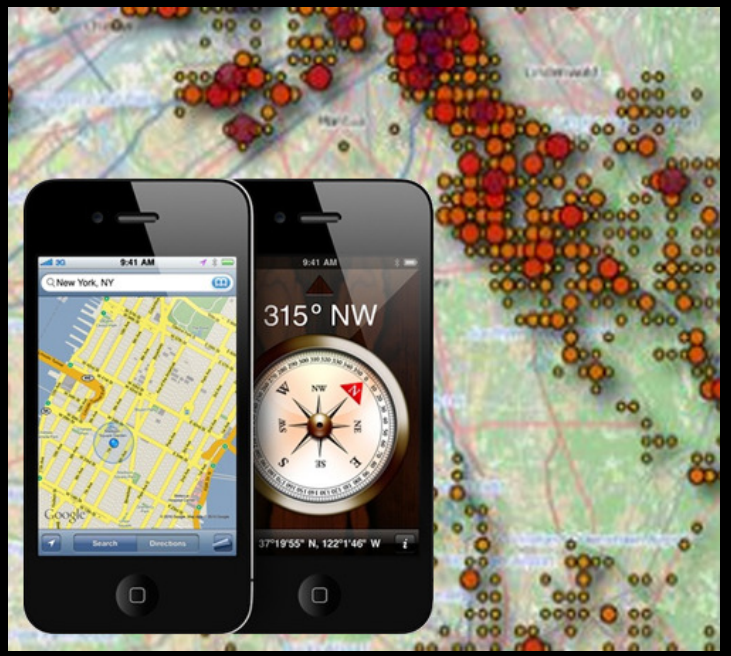

Google


Morgan St


Google


Dislike
Like
German Official: Facebook 'Like' Button Unlawful
Aug 19, 2011 4:13 PM CDT









YOU ARE HERE.
AND WERE THERE,
AND THERE, AND
THERE, AND .....

Outline

- Central question
- Google StreetView case
- Facebook Like Button case
- Apple Location data case
- Comparison and analysis
- Future perspective and conclusion

Central question

- What happened?
- What did companies claim to be the cause:

  Mistake, bug, or intention?

- Main question: What lessons can we learn from the three cases?

Google StreetView

- StreetView cars for making images of streets
- Also collected WiFi data
- SSID and MAC address…
- …and payload data

- Only small bits when driving by

Google StreetView

- 30 GB payload data
- Stored on 714 hard disks

- Contained personal data and sensitive personal data

Google StreetView

- Hamburg DPA asked for audit on collected data
- In a public blog post, Google stated that it was a mistake, because of a piece of code included in the software…
- One of the engineers apparently had enabled this code…
- But it was unintended

Facebook Like Button

- Facebook 

button is displayed on millions of websites
- The button facilitates the tracking of web users by sending information to Facebook
- Members and non-members
- Findings published November 30, 2010:
  - Facebook Tracks and Traces Everyone: Like This!
    - SSRN: http://ssrn.com/abstract=1717563

Facebook Like Button

- Investigations by several parties, including DPA Hamburg and ULD
- Class action lawsuit in California

Facebook Like Button

- No public reaction from Facebook
- In personal communications the practice was denied
- But in response to Hamburg DPA Facebook claimed the cookie setting and tracking to be a ‘bug’

Facebook Like Button

Third, you had questions based upon a reading of a publication you referred to in your letter in which the author described how certain websites that integrated Facebook Connect set cookies that could be read by Facebook and thus Facebook could track non-users' or logged-out users' web activity. What was described in the publication was the result of a bug. Some Connect websites were inadvertently using an old Software Developing Kit (SDK), where a cookie would be set. We fixed the bug as soon as we were made aware of it, and we now run a test every day to ensure that no cookies are being set. We do not have any control over whether third-party websites set their own cookies on visitors' machines, and, if the website does, none of such cookie information would be accessible to Facebook.

Apple Location Data

- File on iPhones and iPads recorded GPS coordinates of nearby Wi-Fi access points and cellphone towers
- Also updated when location services were turned of
- According to Apple, the latter was the result of a bug

Apple Location Data

- Apple claimed not to track individuals
- Location data only in aggregated form to Apple
- Used for more accurate calculation of location of the device
- Future use for navigation services

Apple Location Data

- Intentional
- But not meant/used for tracking
- Nevertheless, the file contained enough data to provide a detailed overview of a user's whereabouts over the past months

Comparison

- In all three cases massive collection of data about individuals
- Different reactions of companies: public blogs or no public reaction
- Admittance or denial
- Intentionally, bug, or mistake in programming

Analysis

Of me:

- Suspicious
- Conspiracy thinker
- Accusive

Analysis

- I prefer ‘awareness raising’

At least I try to be objective based on the available information and…

Analysis

…This leads to some intriguing questions:

- Google: Why was the software code there in the first place?
- Google: How can it be that no one knew about the code being enabled?

Analysis

- Facebook: Why deny the practice or not react at all?
- Facebook: Did they not know or did they not take action?

- Looks like it was not a bug but a feature, and apparently Twitter has recently changed their Tweet button system so they also track users even after logging out

Analysis

- Apple: Aggregated data as sent to Apple no longer personal data, but data on iPhone are. Then, who is processor/controller, and who is responsible if an iPhone is stolen or gets lost?
- Apple: How can data from individual devices be sent to Apple in aggregated form?

Conclusion

- If the data collections were taking place consciously, the companies were not compliant with the DPD
  - Information obligations
  - Purpose specification
  - Legitimate ground for processing

Conclusion

- If the data collections were taking place inadvertently, the companies were not compliant with the DPD
  - Organisational measures
  - Technological measures
  - Unlawful processing

Future perspective

- Major ICT companies are either trying to gradually shift and cross the boundaries of privacy and data protection, or
- They are working with technologies that are so complex that companies cannot appropriately control them

- Both options are very worrisome!

Future perspective

- Business culture should change in order to make prevention of DPD breaches a standard
- Internal processes have to be well-organized and include sufficient control mechanisms

- External checks should become a standard
- DPAs should have the means to do more and better investigations and to enforce the DPD

THANK YOU

Arnold Roosendaal
A.P.C.Roosendaal@tilburguniversity.edu