Privacy vs. Data: Business Models in the digital, mobile Economy

Lecture 7 + 8
Personal Data Collection & Usage

SS 2017

Dr. Andreas Albers

Agenda

- Introduction

- Means of User Identification

- Personal Data Collection

- Personal Data Processing

- Personal Data Usage

- One more thing

Personal data (EU Directive 95/46/EC)

- *'Personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;*

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

3

Steps towards Commercial Personal Data Usage

Personal Data **Collection**

Personal Data **Processing**

Personal Data **Usage**

Requirements for Personal Data Collection

- Availability of a *unique identifier* for an individual (e.g. device ID)

- Access to personal data through
  - observation of the behaviour of individuals (e.g. Behavioural Tracking in the Web)
    → *Implicit/observed Data*
  - deliberate disclosure of personal data through individuals (e.g. posted content on Social Networks)
    → *Explicit/deliberately disclosed Data*

Technological Enablers for Personal Data Collection

- **Online/Mobile Web Logs/Trackers**
  - e.g. Google Analytics, Yahoo Tracker

- **Online/Mobile In-Services/Apps Logging**
  - e.g. creation of log files during service usage

- **OS-Build-in Logging functionality (Desktop/Mobile OS)**
  - e.g. log files generation during service usage

- **Spyware (excluded from course)**
  - e.g. unwanted software secretly installed along a regular software installation process
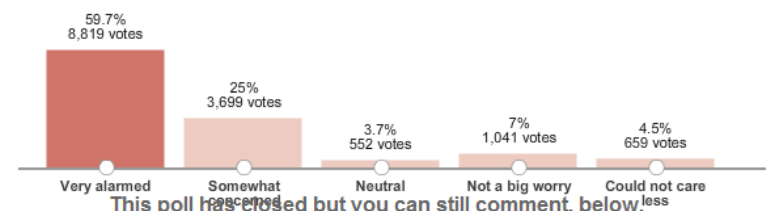
- **First Party (owner/operator of a service or software)**

  - e.g. Amazon tracking individuals visiting products on their website

- **Third Party (organisation unrelated to a service, but enabled to collect personal data) about individuals**

  - e.g. Google Analytics tracking visited news articles on CNN.com

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

7

What „they" (already) know

Site	Exposure Index	Trackers
dictionary.com	Very High	234
merriam-webster.com	High	131
comcast.net	High	151
careerbuilder.com	High	118
photobucket.com	High	127
msn.com	High	207
answers.com	Medium	120
yp.com	Medium	89
msnbc.com	Medium	117
yahoo.com	Medium	106
aol.com	Medium	133
wiki.answers.com	Medium	72
cnn.com	Medium	72
about.com	Medium	83
cnet.com	Medium	81
verizonwireless.com	Medium	90
imdb.com	Medium	55
live.com	Medium	115
att.com	Medium	58
walmart.com	Medium	66
bbc.co.uk	Medium	45
ebay.com	Medium	42
ehow.com	Medium	55
amazon.com	Medium	38

How concerned are you about advertisers and companies tracking your behavior across the Web?

- Very alarmed: 59.7% 8,819 votes
- Somewhat concerned: 25% 3,699 votes
- Neutral: 3.7% 552 votes
- Not a big worry: 7% 1,041 votes
- Could not care less: 4.5% 659 votes

This poll has closed but you can still comment, below.

Source: Wallstreet Journal – What they know, 2014

Some Commercialisation Areas for Personal Data

- **Targeting of Advertisements**

- **Product Recommendations**

- **Service/Content Personalisation**

- **Research**

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

9

Agenda

- Introduction

- Means of User Identification

- Personal Data Collection

- Personal Data Processing

- Personal Data Usage

- One more thing

- Cookies
    - are small text files stored on Desktop PCs and mobile devices
    - have multiple purposes (e.g. session creation, storing of user configuration or personalisation information)
    - and allow user identification

- For *user identification,* a globally unique identifier (GUI) (e.g. number) is stored in the cookie
    - If a user accesses a website, the GUI is collected by the server and referenced to a (maybe) existing user profile stored on the provider's server

- Special types of cookies
    - Flash or Silverlight cookies
    - Used for Flash (Adobe) or Silverlight (Microsoft) applications, but are much more persistent that regular browser cookies
    - Super-Cookie, restores prior delete cookies by users (enabled/operated by Verizon)

- EverCookie (very, very special type of cookie)
    - See for yourself: http://samy.pl/evercookie/ ;-)

- Web Storage (as part of HTML 5)
  - Similar to cookies, but with more storage capabilities and a more sophisticated interface
  - Up to 10 MB storage space per domain

- The storage of (personal) user data is actively supported by the W3C

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

12

- **Internet Protocol (IP) based addressing**
    - Every host and router on the internet has an IP address.
    - An IP address is unambiguous. Two computers cannot use the same (public) IP address at the same time.

- **IPv4 (currently used)**
    - Example: 192.168.133.47
    - IPv4 supports 2^{32} = 4,294,967,296 addresses,
    - But: There are no more unallocated IPv4 internet addresses left!
    - Solution so far: On reconnect to the service provider, a new (dynamic) IP-address is assigned to the client (exception: static IP-addresses)

- **IPv6**
  - An IPv6 address is consists of 128 bits (instead of 32-bit).
  - The new IPv6 address space supports 2^{128} addresses = 340,282,366,920,938,463,463,374,607,431,768,211,456

- **Problem? Let's see …**
  - *"So we could assign an IPV6 address to EVERY ATOM ON THE SURFACE OF THE EARTH, and still have enough addresses left to do another 100+ earths. It isn't remotely likely that we'll run out of IPV6 addresses at any time in the future."* (Steve Leibson)
  - →Every individual could theoretically get a unique IP-address at birth for life.

- **Solution approach towards more privacy for IPv6 addressing**
  - IPv6 "Privacy Extensions" allow randomisation of IP address assignments
  - Problem: Not perfectly working / not activated by default on some OSs
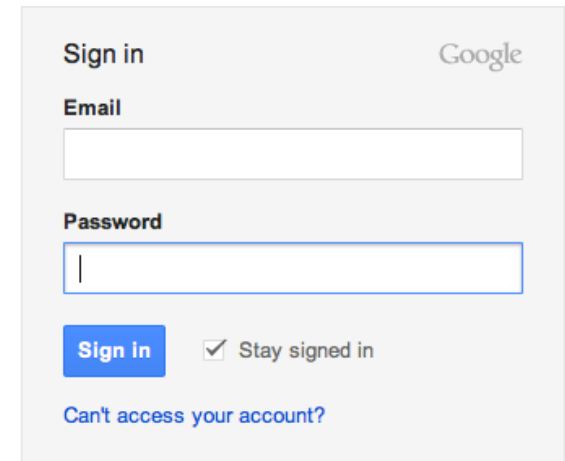
Identifying Individuals Online: Misused Mechanisms

- HTTP/HTML Mechanisms *misused* for user identification
    - Window.Name caching
    - Storing cookies in RGB values of auto-generated, force-cached PNGs
    - Using HTML5 canvas tag to read pixels (cookies) back out (check www.browserleaks.com/canvas to test your browser)
    - HTTP authentication caching
    - …

- ## Sign-On
  - ### Actively logging into Websites (e.g. Google)

- ## Single-Sign-On
  - ### Logging on to multiple sites using a single ID provider (e.g. Facebook Connect)

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

16

Identifying Individuals Online: Browser Fingerprinting

- Identification of users based on numerous available browser information
    - Browser version
    - OS version
    - OS language
    - Screen resolution
    - Installed fonts
    - Installed plug-ins
    - ...

- For more information, visit
    - visit http://browserspy.dk or http://browserleaks.com
    - check your browser fingerprint yourself:
        - https://panopticlick.eff.org/
        - https://www.browserleaks.com/canvas

- ## Mobile Device ID
    - Globally unique proprietary identifier for a mobile device assigned by device manufacturers
    - Accessible for application developers via designated API (Application Programming Interface)
    - Not to be confused with the IMEI (International Mobile Equipment Identity) of the GSM (Global System for Mobile Communications) Standard

- ## Mobile Application ID
    - (Globally) unique proprietary identifier for a mobile apps assigned by application developers or providers
    - Online/Mobile browsers allow only cookies / web storage as means for user identification (which can be deleted at any time by a user)
   - Mobile apps can freely store any kind of information persistently on the mobile device

- Collection of *software* and *hardware* settings acquired from a (mobile) device
  - In their entirety these settings allow a almost unique fingerprint
  - In comparison to the browser fingerprint, a device fingerprint can hardly be changed by a user
    - Changing one setting (e.g. OS version) is often not enough

- Business models based on fingerprinting (example)
  - Blue Cava: bluecava.com
  - Cross-Screen User Identification

bluecava
CROSS-SCREEN STARTS HERE

- Mobile apps and devices allow automatic sign-in without user interaction

- Example
  - Desktop Browser: User has to enter credentials manually and start the sign-in process
  - Mobile App: Credentials are stored within the app and sign-in automatically happens on application launch

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

20
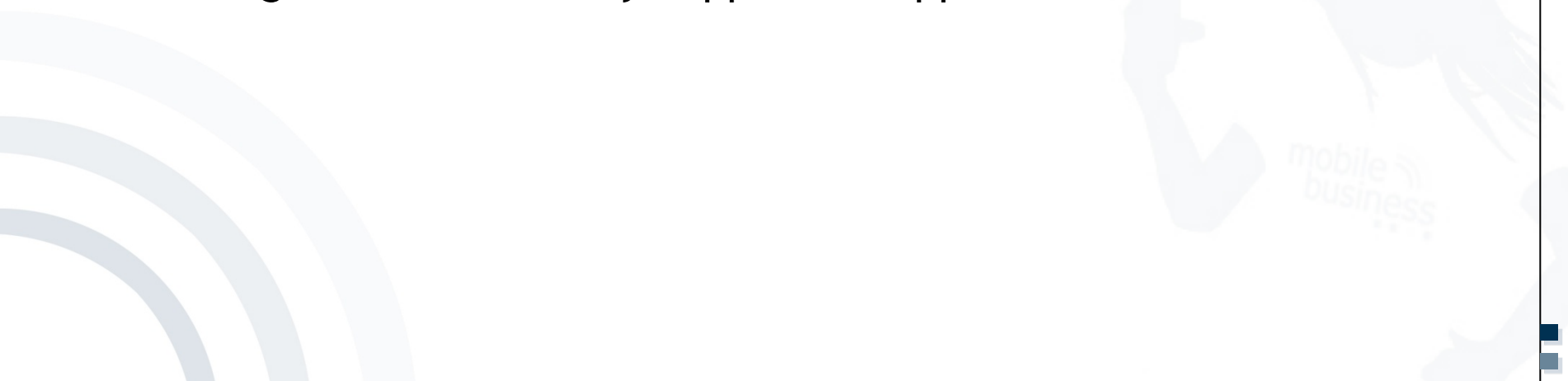
- Similar to browser or device fingerprinting

- Collecting as much as possible information about a user

- Individual information are non-identifying at first (e.g. interests, age, gender, home town, habits, etc.)

- Once there is enough personal data available, the sum of this data (user profile) can be used as identifier for individual users

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

21

Upcoming Means of Identitication

- **Biometric Identification**
  - Face Recognition (already used by Facebook)
  - Voice Recognition (Apple's Siri never forgets a nice chat with its owner ☺)
  - Implanted ID chips
  - Biometric Tattoos
  - Gait recognition
  - Electronic Fingerprint scan
  - Iris Scan
  - …

Source: blog.bufferapp.com

Agenda

- Introduction

- Means of User Identification

- Personal Data Collection

- Personal Data Processing

- Personal Data Usage

- One more thing

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

23

- Online Tracking
- Mobile Tracking
- Mobile-Offline Tracking
- Closed Loop Tracking
- User-generated Personal Data

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

24

Raw Online Tracking Data about Individuals

- **Online Tracking on Websites**
    - User-Accessed websites across different providers (e.g. via Google Analytics)
    - Access time, IP-address, geographic region of access, visiting time, clicked links, viewed content, entered text, submitted requests
    - Browser information (Apple or PC computer, language, OS, etc.)

- **Desktop Tracking Software**
    - Virtually any data, content or user behaviour is theoretically available for trackers

Web/Online Tracking Possibilities

- ▪ **Online Tracking on Websites**
  - ▪ Limited data collection possibilities due to browser as "sandbox" runtime for web applications

  ## VS.

- ▪ **Desktop Tracking Software**
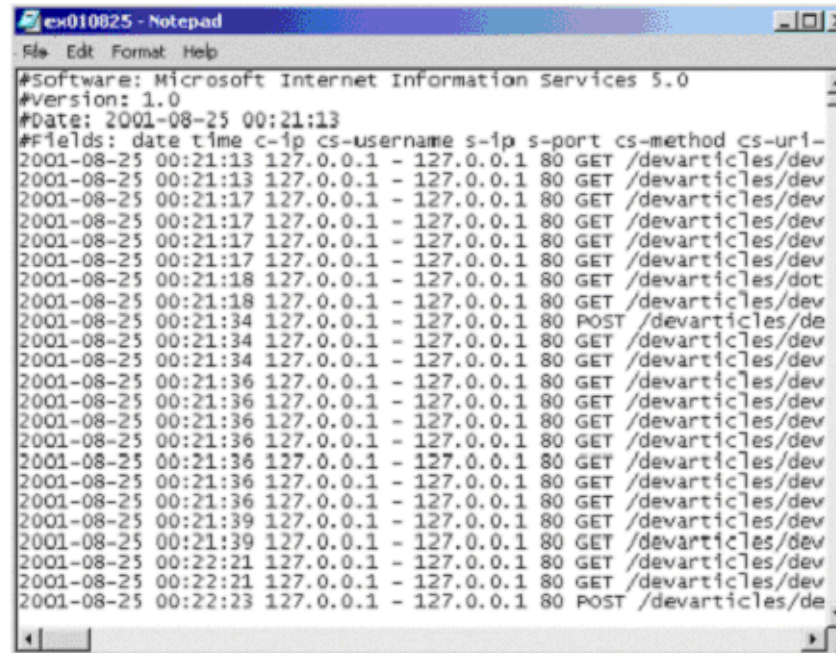 - ▪ Depending to the access and execution rights of the software, there are no limitations for data collection

- ## Web Server Log File
  - A **server log** is a file (or several files) automatically created and maintained by a server about its performed activity.

- ## Log Entry Example
  127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326

- **Browser Information**
  - Browser version,
  - OS Version,
  - OS language,
  - Screen resolution,
  - installed fonts,
  - installed plug-ins,
  - …

- **For more information visit http://browserspy.dk/**

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

28

- **Browser History Privacy Issues**
  - Most web browsers keep record of visited URLs by a users
  - These visited URLs are coloured differently when displayed on a website
  - This colouring can be checked using JavaScript

- **Browser History Stealing**
  - Compilation of a list of websites to check for a user visit (e.g. spiegel.de, tagesschau.de, heise.de, etc.)
  - Asking the browser (one by one) if any of the websites is marked as „visited"
  - NOTE: Thieves cannot ask the browser for a list of visited website. Instead they have to check for each website individually.

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

29

- **JavaScript-based Analytics**
  - Real-time tracking of users visiting a website using JavaScript
  - Analytic services are often provided by Third Parties (e.g. Google) in exchange for access to analytics results for First Parties
  - Client-based approach, which allows a more detailed collection of user behaviour compared to server logs

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

30

- Web Bugs
  - 1x1 pixel size images on websites typically invisible to users and loaded from Third Party servers
  - Allow Third Parties to track page visits of users and place cookies
  - Works in HTML-eMails as well (allows to determine if and when email was read)

- URL Referrers
  - Web servers are able to determine from which prior website a user is coming from
  - Example
    1. User searches on Google for a tech online site
    2. Click on first search results (e.g. arstechnica.com)
    3. Ars Technica receives the URL referrer: www.google.de/search?q=tech site

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

31

- Like/Google+ Button (**without** FB/Google account)
- Websites with Like/Google+ buttons report visitors to Facebook/Google
- Like/Google+ Button works similar to Web Bugs for users without FB/Google account

Website 1

Website n

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

32
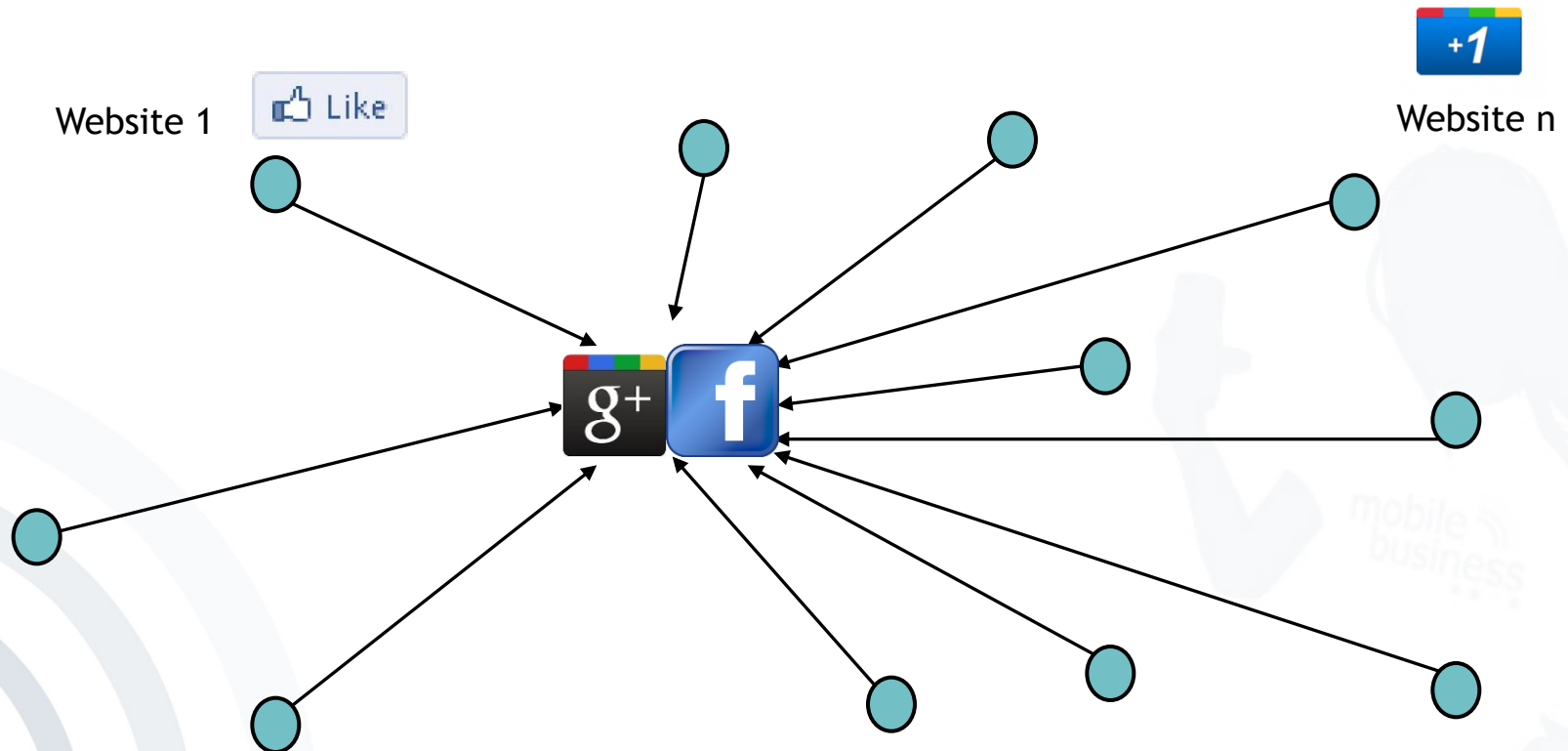
- Like/Google+ Button (user **with** FB/Google account)
- Websites with Like/Google+ buttons report the visited websites of an individual user to Facebook/Google
- This data persistently stored with the account of this user (even if the user is not logged in or actually clicked a Like/Google+ button)

Website 1

Website n
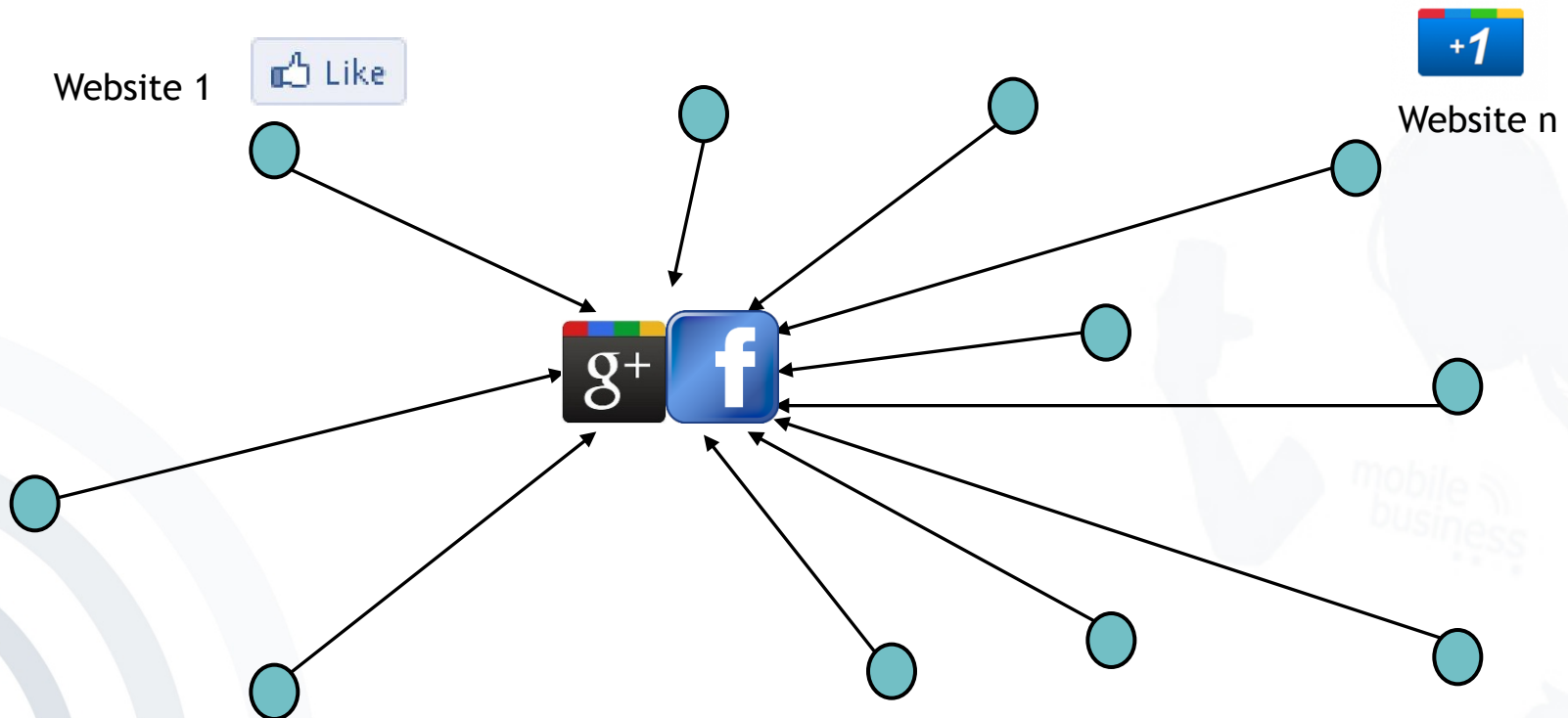
- Facebook Button 2.0 (i.e. Open Graph 2.0)
- Allows to report more detailed user activities – even in real-time (in addition to the existing Like Button)
- Example:
    - Randy watched a Madonna Video on XYZ-Videos.com
    - Jennifer is listing to Bon Jovi on Spotify

Website 1

Website n

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

34

- Facebook Apps

Request for Permission

Chess.com is requesting permission to do the following:

Access my basic information
Includes name, profile picture, gender, networks, user ID, list of friends, and any other information I've made public

Chess.com

By proceeding, you agree to Chess.com's Terms of Service and Privacy Policy and will be taken to apps.facebook.com · Report App

Logged in as Andreas Albers · Log Out

[Allow] [Don't Allow]

- Once access has been granted, Facebook apps have unrestricted access to all or most profile information.

- **Tracking of conducted transactions at Online market places**
  - Apple iTunes
  - Google Android Market
  - Amazon Market Place
  - …

- **Tracking of Online payments**
  - Google Checkout
  - PayPal
  - …

- **Knowledge about confirmed payments for a conducted transaction are more valuable for an advertiser than any knowledge about clicked or viewed online advertisements.**

- World of Warcraft (Example)

- Every activity (e.g. step, click, interaction, fight, communication, etc.) is observed and analytically processed: 24/7 virtual surveillance!
- Can this data be linked to other personal data in the real world?

- **Communication Tracking (e.g. GMail)**
  - Google offers free eMail Service
  - Almost „unlimited" storage space
  - Sophisticated Web Interface

- **What gets Google in exchange?**
  - eMails are stored „forever"
  - eMails are scanned in order to build a user profile
  - Advertisements are displayed and targeted based on existing eMail communications

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

38

- **Cloud-based Data Storage Services provide data about**
  - browser configuration synchronisation activities
  - document synchronisation (e.g. iCloud)
  - cloud-based music & video streaming services
  - cloud-based photo collections
  - used cloud-enabled desktop or mobile cloud applications
  - ….

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

39

Mobile Tracking

- Available Data on mobile websites
  - User-accessed websites across different providers
    (e.g. via Google Analytics)
  - Access time, IP-address, geographic region of access, visiting time, clicked links, viewed content, entered text, submitted requests
  - Browser information (computer type, language, OS, etc.)

- Data collected via mobile device/apps
  - Virtually any data, content or user behaviour is theoretically available (depending on device or OS manufacturer policy/restrictions)
  - Unique, personal data on mobile devices: mobile device/app ID
  - In the near future more sensor data will become available as well
    - e.g. light, noise, temperature, polls, etc.

Mobile Tracking Possibilities

- **Mobile Tracking on Websites**
  - Limited data collection possibilities because browsers or web applications are captured in a "sandbox"

VS.

- **Mobile App/Device Tracking**
  - In general device manufacturers (e.g. Apple) are more restrictive about what kind of apps are allowed on a mobile device (compared to desktop PC applications)
  - However, tracking is allowed by default for mobile app developers
  - Tracking for manufacturers about mobile device and app usage is often "build-in" into the OS

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

42

- **With some limitations, basically the same data sources for mobile web tracking as for online web tracking exist**
  - Web server logs
  - Analytics scripts
  - Browser information
  - History stealing
  - URL referrer
  - …

- **Mobile Device Usage Trackers**
  - Registering mobile device for first time use
  - Tracking of mobile apps launches
  - Tracking of installed applications
  - …

- **Mobile In-App Tracker**
  - Analytics software (First or Third Party) to track user behaviour
  - First Party is typically mobile app provider
  - Third party is typically an advertising network or data provider

- Mobile Apps are increasingly requesting access to personal data (besides access to location data)
    - Contacts, Calendar,
    - Reminders, Photos, …

- Infamous examples:
    - WhatsApp
        - WhatsApp requires access to personal contacts before users can start using the mobile app
    - Facebook
        - Facebook requests/recommends access to e-mail accounts of a user in order to enable friend finder features (based on e-mail addresses)

- Facebook / WhatsApp
    - What was Facebook looking for when it bought WhatsApp? Personal Data?
    - Why were so many WhatsApp users threatening to leave WhatsApp?

- Location Tracking

  - Tracking of the different locations every time a user launches a location-based service app on his mobile device

- Localisation Technologies

  - WiFi (the mostly frequently used)

  - (Assisted)-GPS

  - Mobile Network: COO (Cell of Origin)

  - Bluetooth, IrDA, NFC

  - Beacons

  - Barcode Scanning

  - Manual user entry

  - Sound-based localisation

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

46
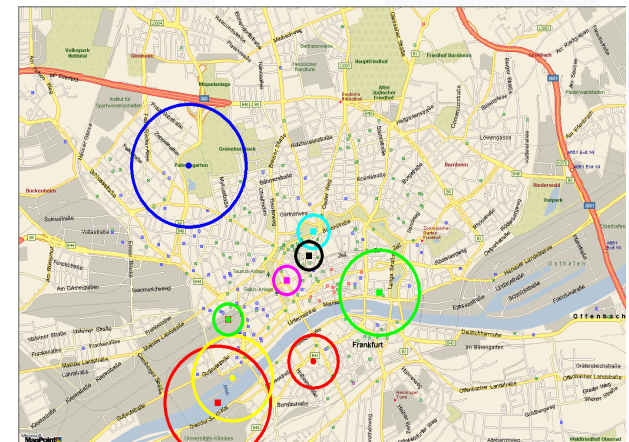
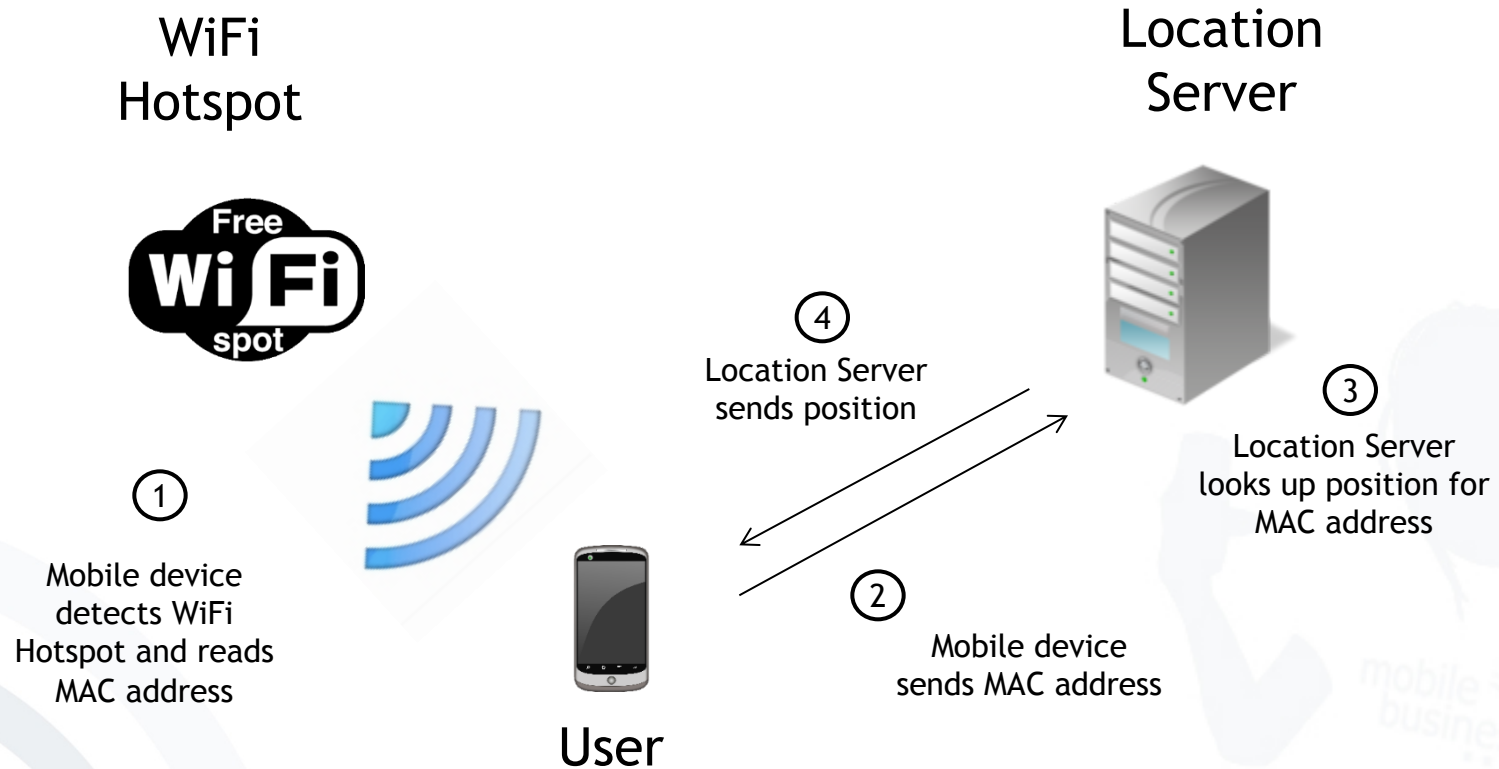# Preparation

- WiFi Hotspots in a city are collected and stored in a central database

  - Collection via plane or by car (e.g. Google Street View cars)

  - WiFi Hotspot data record contains geo-position and MAC-Address of WiFi Router

  - Two location providers world-wide so far

    - Google and Skyhook Wireless

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

47
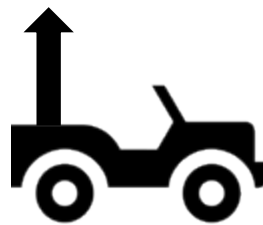
WiFi Localisation Process

WiFi
Hotspot

Location
Server

④
Location Server
sends position

③
Location Server
looks up position for
MAC address

①
Mobile device
detects WiFi
Hotspot and reads
MAC address

②
Mobile device
sends MAC address

User

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

48

- **Advantages**
  - Quite exact: approx. 20m
  - Very fast localisation
  - WiFi available in many mobile devices

- **Disadvantages**
  - Privacy issues due to location submitted to location providers during positioning process
  - Only two location providers currently exist (location profiles about individuals become possible)
  - Increased power consumption in non-stop operation.

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

49

GPS Localisation

- GPS – Global Position System is operated by US-Department of Defence (USA)

- Four satellites needed for exact positioning (X,Y,Z + Time)

- Assisted GPS calculates the rough location based on Cell-of-Orgin first in order to then speed up the GPS based positioning process

- Other GPS systems
  - Galileo (European Union)
  - GLONASS (Russian Space Forces)
  - Beidou/Compass (China)

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

50
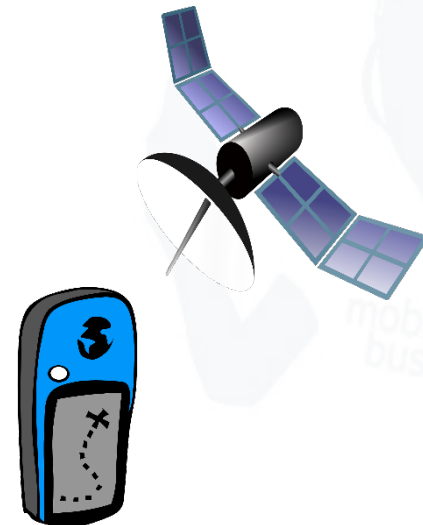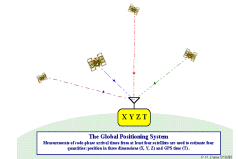
- **Advantages**
  - Quite exact: 5-15m
  - Low cost chip sets, embeddable in terminals
  - Large choice of standard software for applications available.

- **Disadvantages**
  - Works only outdoors
  - USA can manipulate or disconnect the signals whenever they want
  - Long initialisation time (up to 3 minutes), which can be improved by "Assisted GPS"
  - High power consumption in non-stop operation.

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

51

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

52
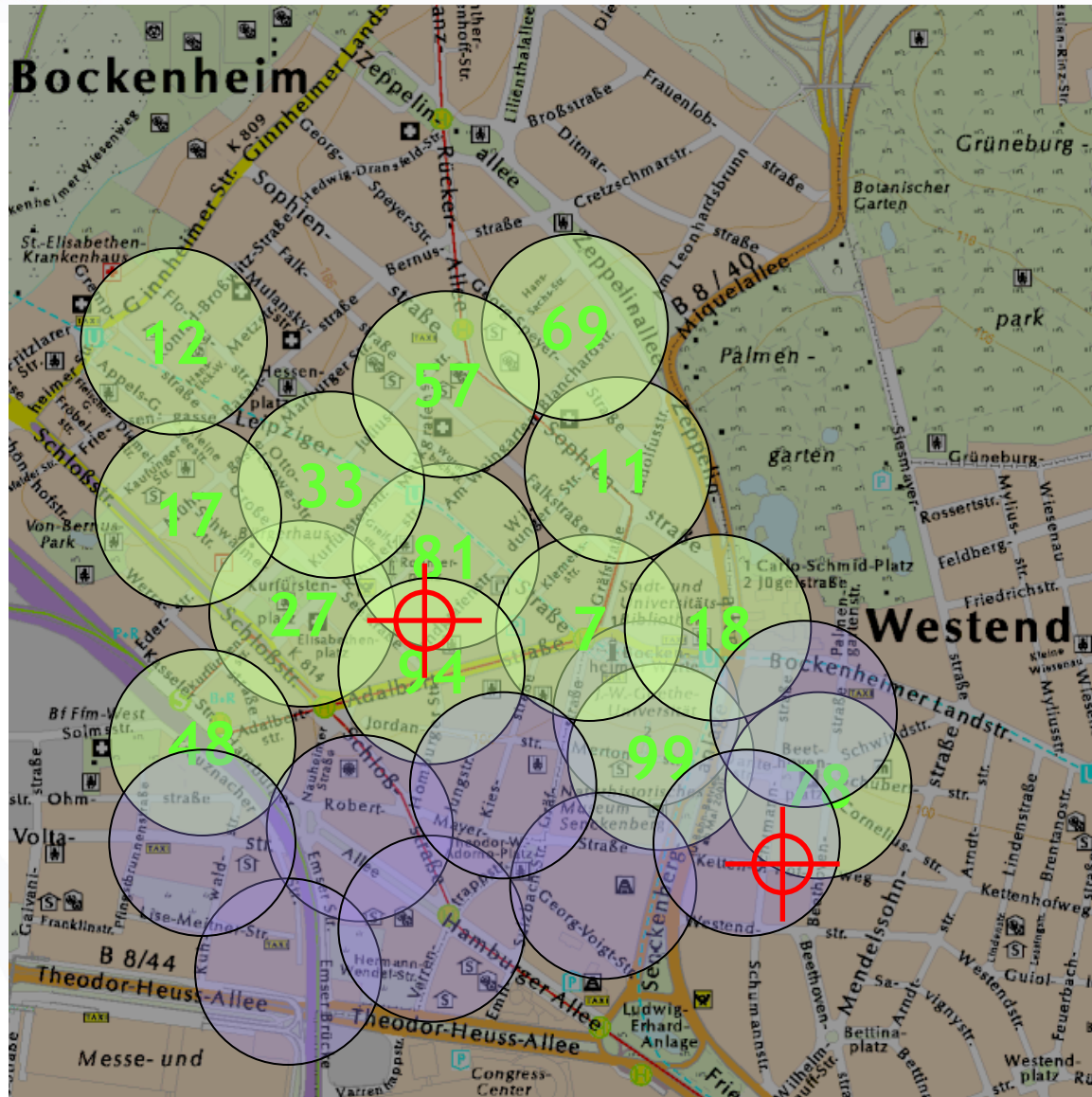
- ## COO Advantages
  - Works with every mobile device because it is mobile network-based
  - Very fast localisation

- ## COO Disadvantages
  - Accuracy depends in the size of the cell
  - 300m in city centers, up to 30km in rural areas
  - Privacy issues due to the fact that the mobile operator always knows the location of the user
  - Costs may occur for localisation requests

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

53

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

54
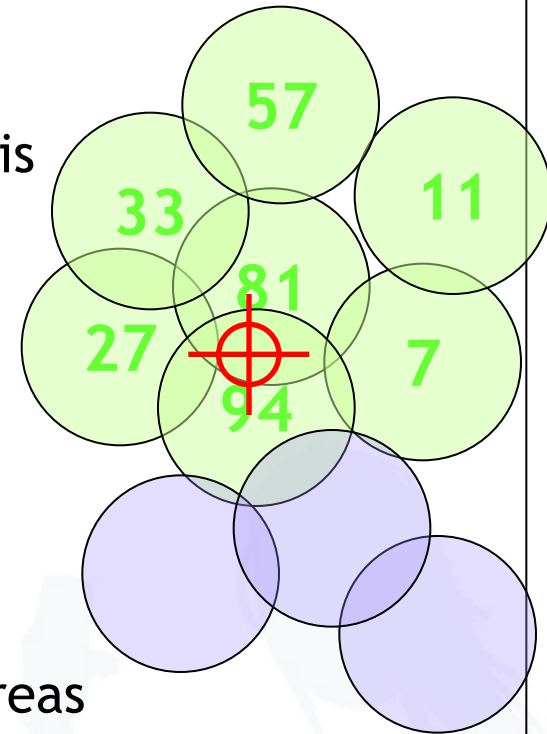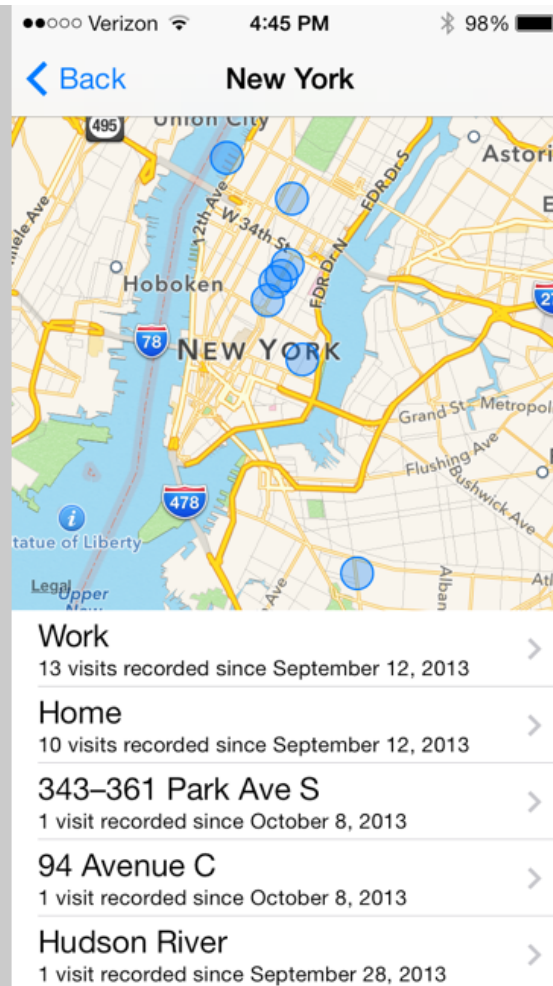
- iBeacons are low-power transmitters (leveraging Bluetooth Low Energy technology), which are used to send push messages / location data to mobile devices in their close proximity.

Source: Forbes.com, 2014
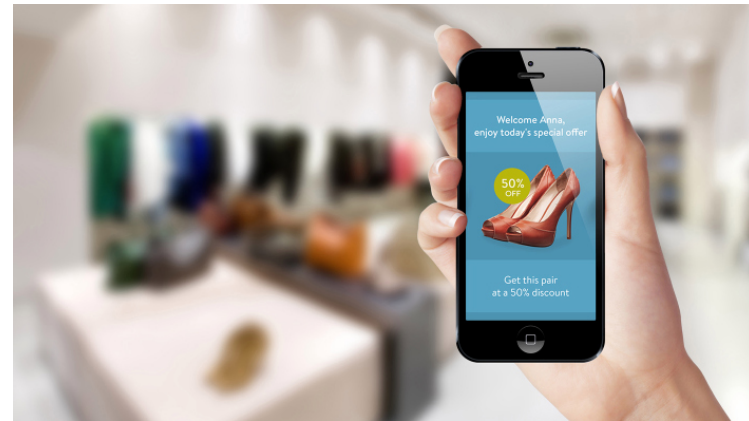
Source: 9to5Mac, 2014

- Allows even retargeting of user from the mobile to the online world.

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

55

Cross-Channel Data Collection
Online – Mobile - Offline

1. At lunch time, Frank searches on his office desktop PC via **Google Search** for "tasty sandwiches".

2. **Google Search** personalises his search results based on his search history and displays **Google AdWords ads** based on Frank's search query as well as his Google+ profile.

3. Frank is displayed four **Google AdWords** ads on the Google search result page and Frank clicks on the second ad "Subways".

4. Frank is transferred to the ad's landing page (i.e. Subways website). There, **Google Analytics** checks what kind of food Frank is interested in – by observing his browsing behaviour.

5. Based on Frank's derived tastes, **Google Offers** sends him a coupon for a Subway store close to his office directly on his phone. It is stored in the **Google Wallet** application of his phone.

6. Frank goes to Subways, pays his lunch and redeems his coupon both via the **Google Wallet** application. He receives a discount in exchange from Subways on his ordered food.

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

57

- **Benefits for Frank**
  - Highly personalised support service for satifying his needs while receiving a discount

- **Involved Google Services**
  - Google Search, Google+, Google AdWords, Google Analytics, Google Offers, Google Wallet

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

58

- **Benefits for Google**
  - Closed Loop Marketing

- **Google knows**
  - how Franks search query relates to the clicked advertisement (Google Search)
  - how Frank interacts with advertisers (Google Analytics)
  - that Frank actually bought a product, which was advertised via Google at the store (Google Wallet & Offers)

- **The Closed Loop**
  - Information about every phase of Frank's transaction is eventually fed back into Google Search and Google AdWords in order to improve the future personalisation of the service and in particular the targeting of the ads.

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

59
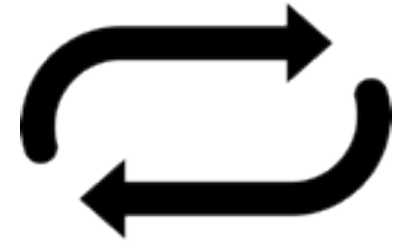
- # Internet of Things

  - Increasingly "things" rather than "people" get connected to the Internet

  - Example: The in(famous) refrigerator, which orders food from the grocery store whenever its about to run out of the latter

    - More serious examples, please …

- Using Internet-enabled devices to manage and check home appliances such as

    - Heating

    - Windows shutters

    - Lights in the house/apartment

    - Locking of doors

    - Devices (e.g. TV, washing machine, etc.)

    - …

- Problem: This could enable your service provider to get insights about what your are doing at home

    - No problem?

winfwiki.wi-fom.de

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

61
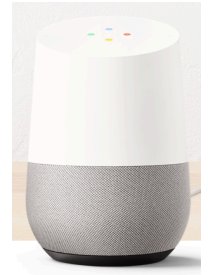
- SmartTV are TVs, which can be connected to the Internet in order to allow
    - on-demand streaming of movies/tv-shows
    - access to the web (browsing, video/ audio streaming)
    - powering TV-apps like twitter, tv-guides, games, etc.
- Privacy Issues
    - Same as Online and Mobile BUT:
    - TVs are also calling secretly „home" in order to share with their manufacturers what his owners are currently watching.
    - Further, TVs with voice control may "accidently" overhear your conversations in the living room.

- A **smart meter** is usually an electrical meter that records consumption of electric energy in intervals of an hour or less and communicates that information at least daily back to the utility for monitoring and billing purposes

Source: Federal Energy Regulatory Commission Assessment of Demand Response & Advanced Metering (2011)

- Advantages
  - No more service people at home required in order to check the current reading of a meter– everything can be done remotely
  - A user can also check the power consumption of every single device in their household (if connected to a single power point)

- Disadvantages
  - Service Providers could be able to derive certain information about users, which may be considered private, e.g.
    - what kind of devices a user owns
    - when a user is at home
    - how many people are currently in the house/apartment
    - in which room of his house/apartment a user currently is
    - what movie is currently being watched (based on the power consumption pattern of his TV)

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

64

- Miniature electronic devices that are worn by the bearer under, with or on top of clothing

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

65

- **Display integrated into glasses**
  (like head-up displays in cars or planes)

- **Context-dependent online services**
  - Based on user location, heading, contents in sight of user, etc.
  - For instance, indoor directions, face or sign recognition, context-based reminders, etc.

- **Personal data Usage**
  - All collected context data (incl. video stream) has to be transferred to Google for processing
  - Users could secretly take photos or videotape their proximity

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

66

Other „Listening" Devices

- Mainly Voice Activated Devices
  - Xbox One
  - Apple Siri
  - OK Google
  - Amazon Echo
  - Barbie
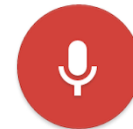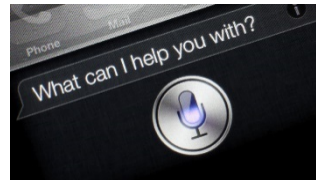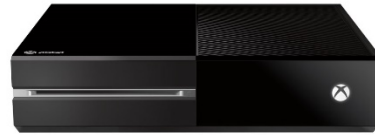
- Due to the complexity of voice recognition, the command is recorded and then send to the cloud for analysis.
- For voice activation via Command („Hey Siri" or "OK Google") has the device has to listen to ALL user commands first (hopefully only locally on the device).

User-generated Personal Data

User-generated Personal Data

- Typical Categories in Social Media
    - Links
    - Pictures
    - Videos
    - Status updates
    - Comments about other individuals
    - Profile information
    - Messages

User-generated Personal Data

- **Additional "Secondary" Data**

  - **Social graph**
    - Relationships to other individuals

  - **User Groups**
    - Recipient groups for user-generated personal data (content)

  - **Privacy Settings**
    - Privacy settings towards other individuals or groups of individuals

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

70

- **Daily Roads**

  - Enabling continous video streams from vehicle via a mobile app

  - If successful, it's Google Street View „on steroids"
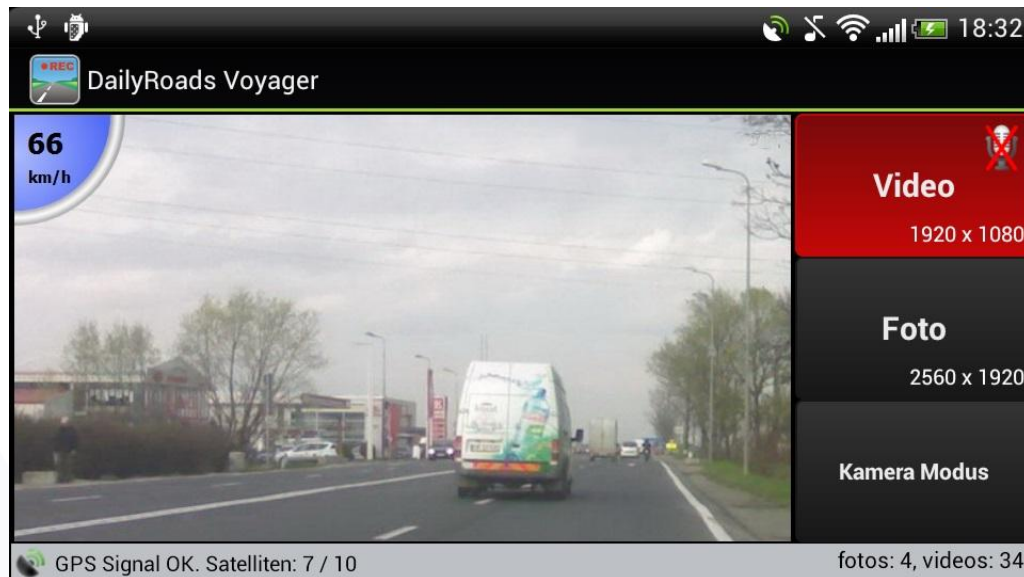
Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

71

Agenda

- Introduction

- Means of User Identification

- Personal Data Collection

- Personal Data Processing

- Personal Data Usage

- One more thing

Example: Facebook Ad Profiling

Deine Einstellungen für Werbeanzeigen

Wir zeigen dir Werbeanzeigen basierend auf den Dingen, die dir wichtig sind. Deine Einstellungen umfassen Informationen aus deinem Profil sowie Handlungen auf und außerhalb von Facebook. Füge Einstellungen hinzu oder entferne sie, damit dir relevante Werbeanzeigen angezeigt werden. Mehr dazu.

Füge eine Einstellung hinzu	Durchsuchen

▶ **AUSBILDUNG** 14

▶ **EINKAUFEN UND MODE** 7

▶ **ESSEN UND TRINKEN** 7

▶ **FITNESS UND WELLNESS** 1

▶ **GEWERBE UND BRANCHEN** 65

▶ **HOBBYS UND AKTIVITÄTEN** 30

▶ **LIFESTYLE UND KULTUR** 24

▶ **NEUIGKEITEN UND UNTERHALTUNG** 177

▶ **PERSONEN** 44

▶ **REISEN, ORTE UND VERANSTALTUNGEN** 16

▶ **SPORT UND OUTDOOR** 3

▶ **TECHNOLOGIE** 37

▼ **REISEN, ORTE UND VERANSTALTUNGEN** 16

tripadvisor Travel/Leisure

Schottland

Rheinland-Pfalz Wahrzeichen

Wiesbaden Stadt

Taunus Outdoor

Mainz, Germany Stadt

Mainz Stadt

Zeit

BBC Travel Local/Travel Website

Hochtaunuskreis Bezirk

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

73

- Google+ / Facebook knows about
  - all posted content (e.g. profile information, posted pictures, comments, etc.) and
  - all "liked/+1" content
  → **User-generated (explicit) personal data**

- Google+ / Facebook knows about
  - (many) visited websites through Like/+1 button
  → **Tracked (implicit) personal data**

→ Towards Predictive Behavioral Targeting

From raw personal data records can be derived ...

- Geographic movement profiles (from tracked locations)
- Personal users interests (from visited websites)
- Education (from visited websites)
- Spoken language (from browser / devices information)
- Used technology to access the Internet (from OS information)
- Estimation of age and gender (from viewed online content)
- Online/offline shopping habits (from transaction history)
- Mobile & online & offline habits (mobile wallet as link)
- Relationships between users (from social networks)

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

75

External Data Cross-Referencing

- Cross-referencing of individual user profiles with

  - Socio-demographic data
    - Results in statistical information about age, gender, education, income, occupation, geographic area of an individual

  - Personal social data
    - Results in statistical information about general living habits, related persons, degree of relation, relation between persons of an individual

  - Business data
    - Results in in statistical information about partnerships, supervisor-employee relationship, education, etc. about an individual

- Foundation of cross-referencing tracking results are commercial statistical data providers such as of Nielsen & Co.

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

76

- Introduction

- Means of User Identification

- Personal Data Collection

- Personal Data Processing

- Personal Data Usage

- One more thing

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

77

- # Explicit "Personal Data"
  - is explicitly expressed by individuals
  - can deliberately be wrong
  - can be misunderstood
  - can be outdated
  - can be incomplete
    - on purpose
    - without an individual knowing

- # Implicit "Personal Data"
  - is derived from the observation of individuals' behavior
  - can be derived wrong
  - can be interpreted wrong

mobile business
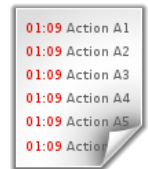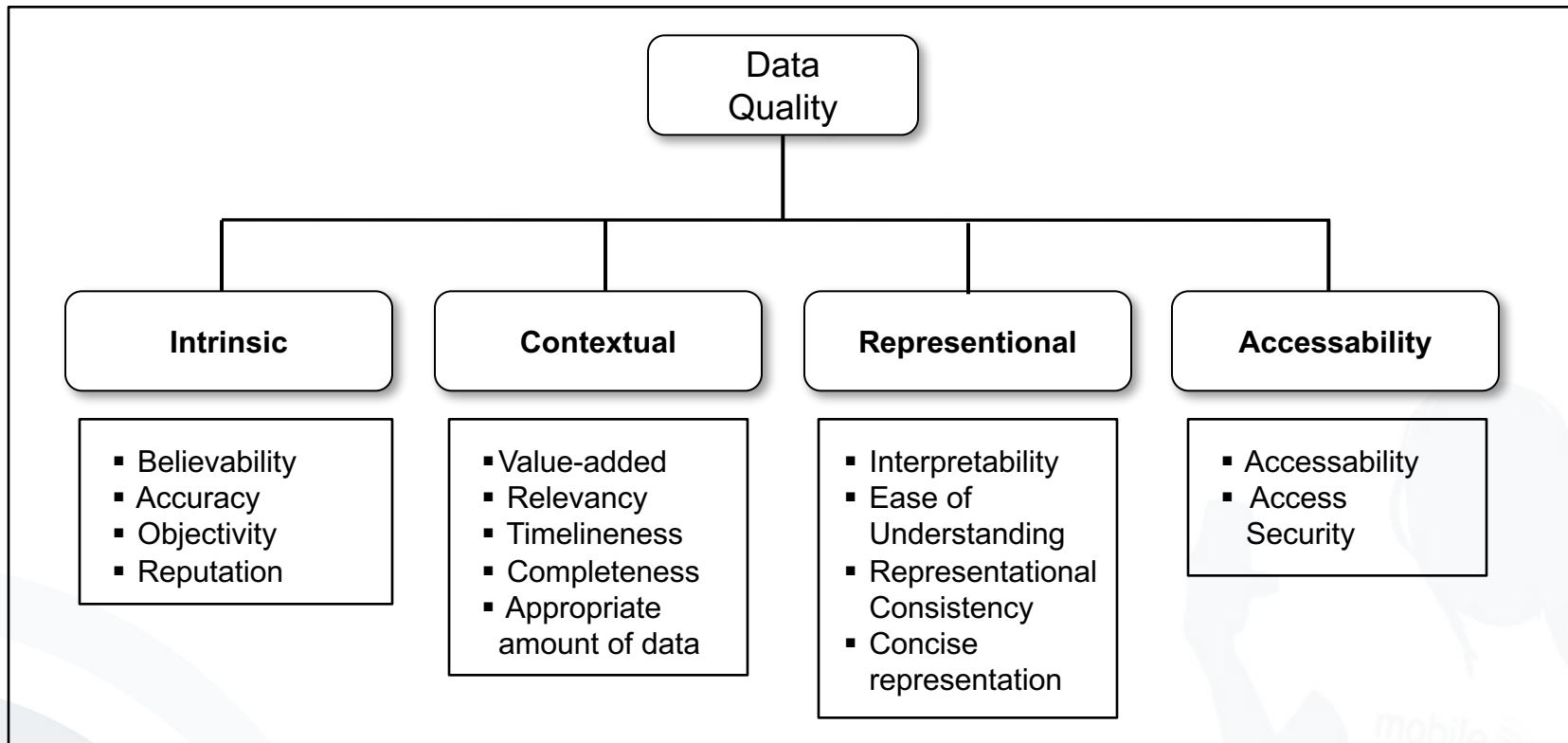
Source: Wang and Strong (1996)

Data Quality

| Intrinsic	Contextual	Representional	Accessability
▪ Believability ▪ Accuracy ▪ Objectivity ▪ Reputation	▪ Value-added ▪ Relevancy ▪ Timelineness ▪ Completeness ▪ Appropriate amount of data	▪ Interpretability ▪ Ease of Understanding ▪ Representational Consistency ▪ Concise representation	▪ Accessability ▪ Access Security

Personal Data Usage Application

- Product Recommendations

- Personalisation of Content & Services

- Targeting of Advertisements

- Price Discrimination

- (Market) Research

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

80

Example Amazon's Value Creation

Product Recommendation drives „Convenience" and „Customer Experience"

Frequently Bought Together

Customers buy this book with Steve Jobs by Walter Isaacson   Hardcover $17.88
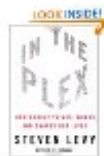
Price For Both: $31.86

[Add both to Cart]   [Add both to Wish List]

One of these items ships sooner than the other. Show details

Customers Who Bought This Item Also Bought

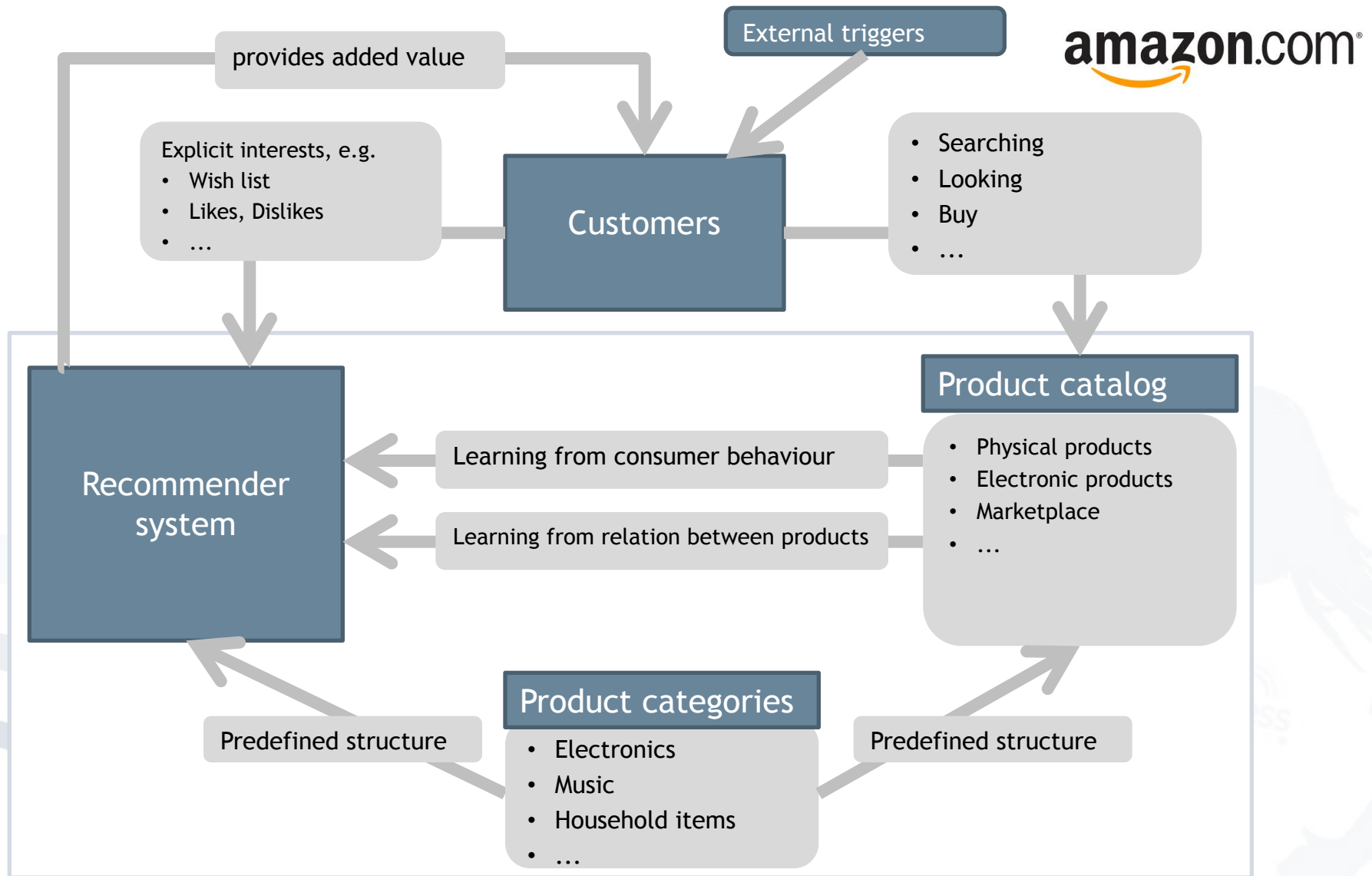I'm Feeling Lucky: The Confessions of Goog... by Douglas Edwards
★★★★☆ (55)
$15.84

The Facebook Effect: The Inside Story of the... by David Kirkpatrick
★★★★☆ (71)
$6.40

The Wikipedia Revolution: How a Bunch of Nobodies C... by Andrew Lih
★★★★☆ (16)
$10.00

mobile business

amazon.com

External triggers

provides added value

Explicit interests, e.g.
- Wish list
- Likes, Dislikes
- …

Customers

- Searching
- Looking
- Buy
- …

Recommender system

Learning from consumer behaviour

Learning from relation between products

Product catalog
- Physical products
- Electronic products
- Marketplace
- …

Predefined structure

Product categories
- Electronics
- Music
- Household items
- …

Predefined structure

Personalisation based on explicit User Preferences

- **Technical Targeting**
  - Using of, among others, IP address, browser model, or operating system for targeting purposes

- **Contextual Targeting**
  - Mainly text-based contents of an accessed website are analysed and Online Marketing campaigns are targeted accordingly.

- **Keyword Targeting**
  - According to the keyword or query provided by an online user, the respective Online Marketing campaigns are displayed.

- **Behavioural Targeting**
  - This approach observes the behaviour of users. It attempts to derive their potential needs statistically as a targeting foundation from the acquired *click stream*.

- **Profile-based Targeting**
  - Profile-based targeting requires that online users voluntarily disclose personal information (e.g. age, gender, personal interests).

A simple "theoretical" example …

- **Assumption:** Individuals, who own an Apple computer have in general a higher willingness to pay for certain kinds of products
- Internet Browsers transfer the OS version of a client to the web server of a merchant (MacOS or Windows)
- Possible resulting price discrimination on the merchant website

PC User: 450€
      Mac User: 499€

Another simple "practical" example …

- Merchants, among other criteria, rate your credibility based on the geographical area you live in (using postal codes)
- If you live in a "bad" area, you most like won't get offered to pay via invoice or direct debit. Instead you have to pay cash in advance.

- **Example: Apple Research Kit**
  - Collects health data from volunteers using the sensors of attached to iPhones

  - Requires explicit user consent

- **Other option to use data for research**

  - Anonymize formerly personalised data
  - But how to anonymize data?

- Unfair price discrimination
- Unintended/unfair reviews (e.g. credit worthiness / health reports)
- Spamming or Fraud
- Sales of contact information to Third Parties
- Cyber-Stalking
- Unintended linking of social roles
- Black Mailing

Agenda

- Introduction

- Means of User Identification

- Personal Data Collection

- Personal Data Processing

- Personal Data Usage

- One more thing

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

88

- Ocean Method to derive the Big Five characteristics of a person with regard to …
    - Openness, Conscientiousness, Extraversion, Good Nature and Neuroticism
    - Allows to conclude on the needs and fears of a person (basically on a persons tendency to behave)
- For this Facebook simple likes are enough …
    - 68 Facebooks likes are enough to determine skin colour, homosexuality, Democrat vs. Republican of a person
        - Further: Intelligence, religion, alcohol, cigarettes and drugs abuse
    - 70 likes to better know a person than his colleagues
    - 150 likes to better know a person than his parents
    - 300 likes to better know a person than his partner
    - 300+ likes to better know a person than her knows himself
- What to do with the following information about a person?*

_____

* https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/

Internet of Things & Data Collection

- http://techcrunch.com/2015/01/25/what-happens-to-privacy-when-the-internet-is-in-everything/
- http://arstechnica.com/security/2016/01/how-to-search-the-internet-of-things-for-photos-of-sleeping-babies/
- http://www.heise.de/ct/artikel/Sengled-Pulse-LED-Birnen-mit-Lautsprecher-2282174.html
- http://www.heise.de/newsticker/meldung/Datenschutz-Werbe-Tracker-ueberwinden-Geraetegrenzen-2921817.html
- https://netzoekonom.de/2015/12/20/mit-daten-von-40-mio-handy-nutzern-verkehrsstroeme-exakt-messen/

Last access: April 2017

Privacy vs. Data, SS 2017, Chair of Mobile Business & Multilateral Security

90