Protecting User Privacy in the Age of Big Data

NAE German-American Frontiers of Engineering Symposium

April 2015

Big-Data: Acquisition, Storing, Analysis, …

RFID

Pictures, voice, gesture …

Quelle: www.bmw.de

Quelle: ww.samsung.com

Power supply

GPS

Medical Data

Meta data, connection data, user profiles

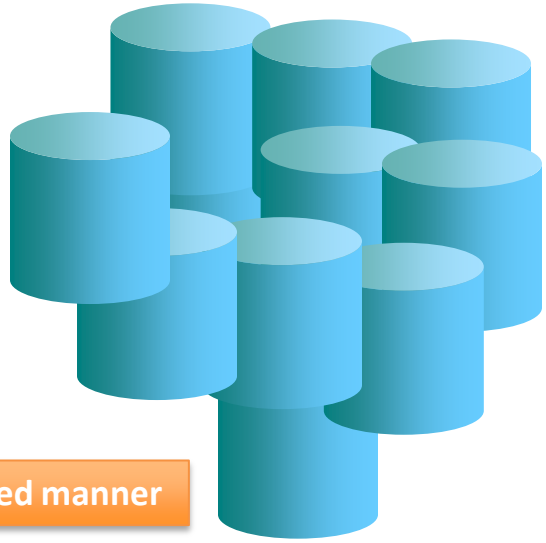
Dauerhafte Verbindung

Dauerhafte Verbindung

Huge amount of data can be efficiently processed

Huge number of places for storing personal data

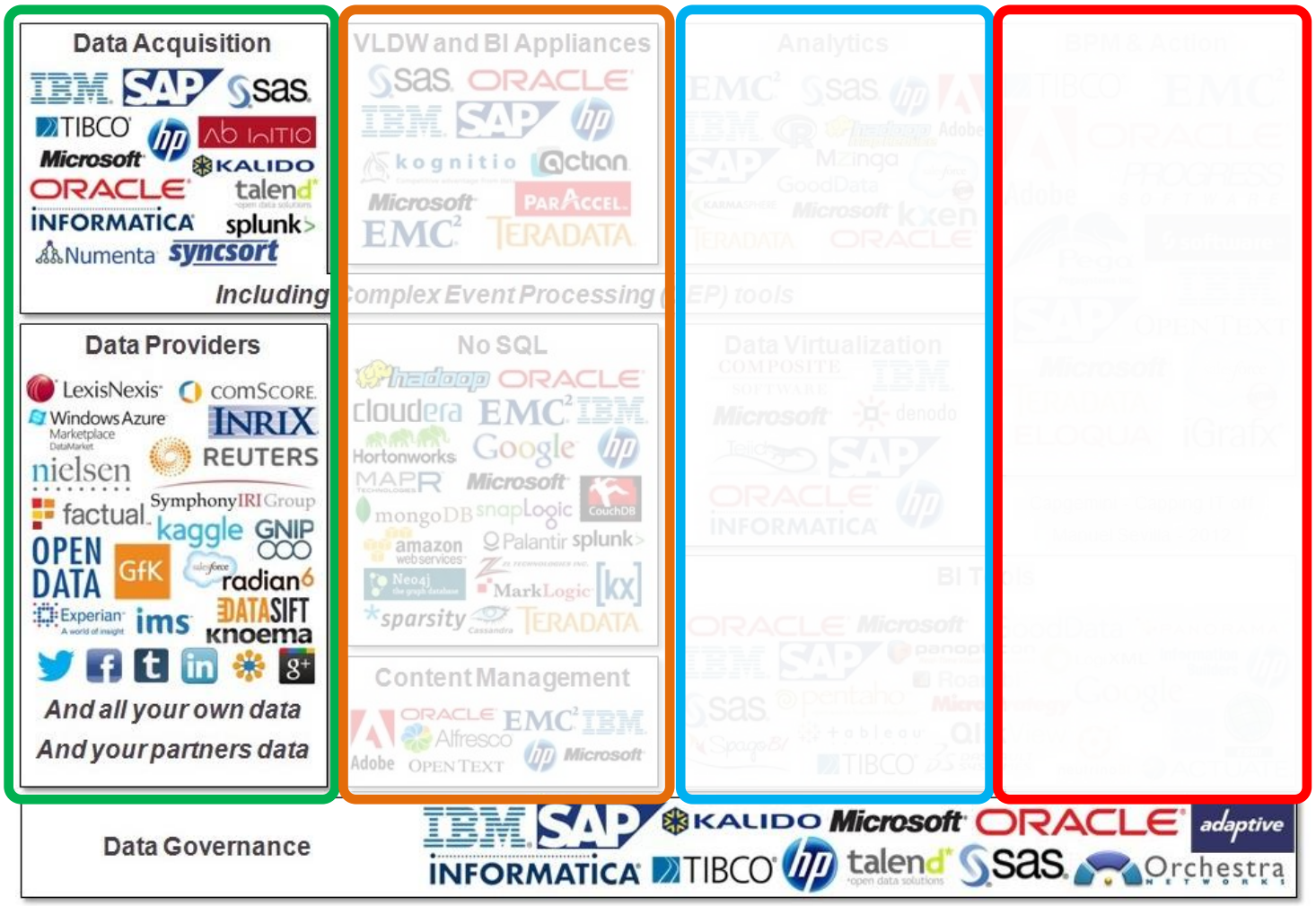
Permanent storage of extremely large datasets



Huge Number of different data sources

Companies collect personal information in an unprecedented manner

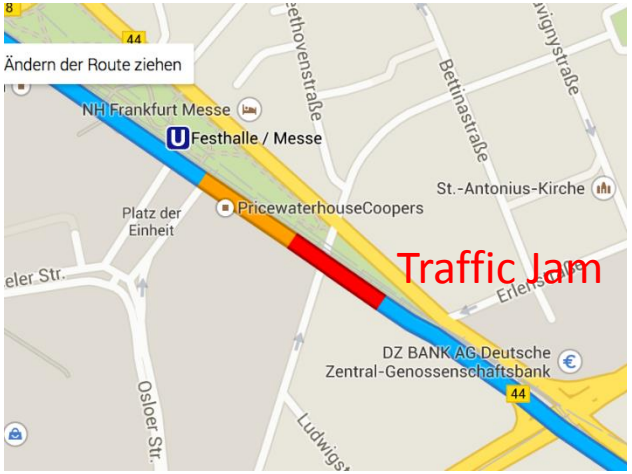
Big-Data: Background and business models



Source: Capgemini, Manuel Sevilla

The Promise of Big Data

- Many advantages because of Big Data:
  - Global knowledge about traffic, and traffic jam prevention
  - Search queries in the Internet (Google, Bing, etc.)
  - Personalized online advertisements
  - Social contact management, self determination (social networks, etc.)



- Many advantages also for research
  - Medicine (disease diagnosis, therapies, DNA analysis)
  - Computer Science as the driving force for efficient analysis
- Big-Data can help sharpen predictions
  - Influenca predictions: http://www.google.org/flutrends/)
  - Improved weather forecast
  - Crime prediction
    http://www.govtech.com/public-safety/Predicting-Crime-Using-Analytics-and-Big-Data.html

Big Data Also Threatens User Privacy Adversaries with Different Motivations

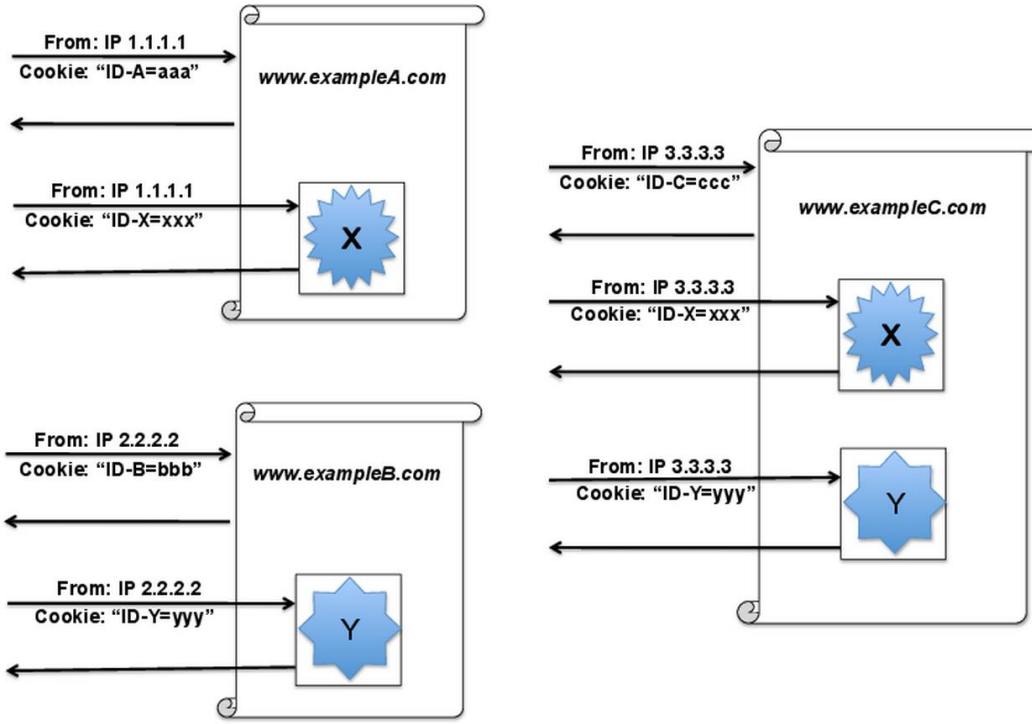
- **Social:** User tracking and inference
  - De-anonymization
  - …often to the surprise of user expectations

- **Commercial:** Choosing not to offer insurance, sell house, etc. based on information gleaned from search, other sources ("redlining").

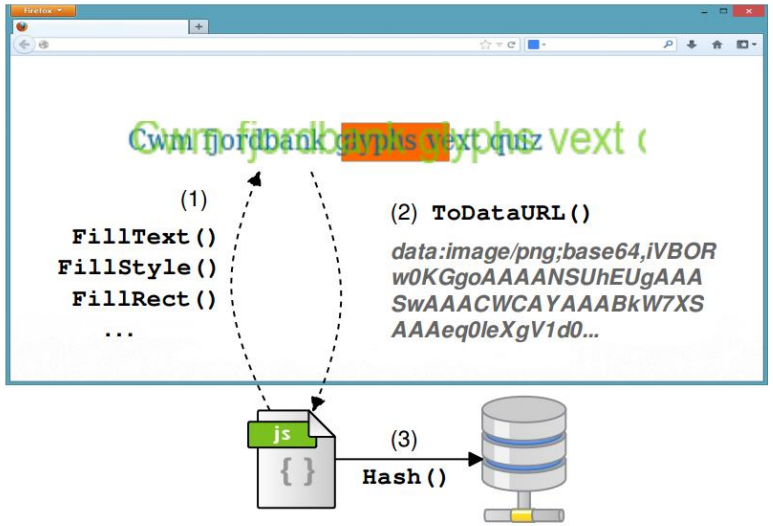
- **Political** attacks, threats, and repercussions

Data Collection and User Privacy

Web Cookies



Canvas Fingerprinting



Many other ways: Flash cookies, "Evercookies", device fingerprinting, etc.

SSL/Encryption is not automatic protection!

Implications for Surveillance

- Example: NSA
  - has authority to collect/observe US-originating or US-bound traffic

- A user may browse a local website, but a large fraction of third-party trackers are in the US
  - 13% from traffic from European users, 20% of traffic from Asia can be linked across multiple sites

Englehardt *et al.* "Cookies That Give You Away: Surveillance Implications of Web Tracking", *WWW* 2015.

We Must Build and Systems That Operate in Adversarial Environments

(and, the average user must be able to use them)

Much Work to Do

- Theory: Understanding What to Build
  - Differential privacy
  - Formal models of anonymity
- Practice: Building and Deploying the Systems
  - (Well designed) tools for circumventing tracking and surveillance
  - Tools and interfaces for increasing user awareness

Speaker Agenda

- Understanding threats
  - Politically motivated adversaries
    Phillipa Gill (Stony Brook University)
- Developing Defenses
  - Improving user awareness
    Matthew Smith (University of Bonn)
  - Privacy-preserving data collection
    Matteo Maffei (Saarland University)
  - A better “software stack”
    Roger Dingledine (The Tor Project)