CHALLENGES IN A HIGHLY INTERCONNECTED DIGITAL WORLD: THE CASES OF PRIVACY AND DIGITAL BIAS

July 2017

Giorgos Flouris (fgeo@ics.forth.gr)

Introduction

- My research
  - Artificial Intelligence
  - Data and Knowledge Representation
    - Without ethical/legal extensions

- This talk deals with two challenges
  - Privacy
    - CAPrice project (https://www.caprice-community.net/)
  - Digital bias
    - No specific results yet

CAPrice

collective awareness platform for privacy concerns and expectations

3

Privacy

Current situation, and the CAPrice project

Joint work with:
Theodore Patkos, Panagiotis Papadakos, Ioannis Chrysakis



Giorgos Flouris, July 2017

Privacy and new technologies

- New technologies pose new privacy challenges
  - Improved ability to gather data
    - IoT, connected cars, smart phones, watches, homes, TVs, baby monitors, …
  - Improved ability to process data
    - Big data, machine learning, deep learning, data science
    - Predict epidemics, personalized medicine, smart cities, …
  - Data-centric business model
    - Personal data: currency for buying "free" products/services
  - Fast technological evolution
    - Moving target, too fast for legislators

Reinstating privacy

- Counter-measures
  - Several technical solutions
  - Many initiatives
  - Legislation (GDPR included)
- Limited effectiveness
  - Reason #1: People don't care
  - Reason #2: Can't fight the big companies
  - Reason #3: Inappropriate legal system
  - We think otherwise:
    Lack of awareness/education

#1: People don't care?

- A common misconception
  - Due to lack of awareness
- People don't understand the risks
  - What data is being gathered
  - What can be done with this data
  - Unaware of related mitigating technical solutions

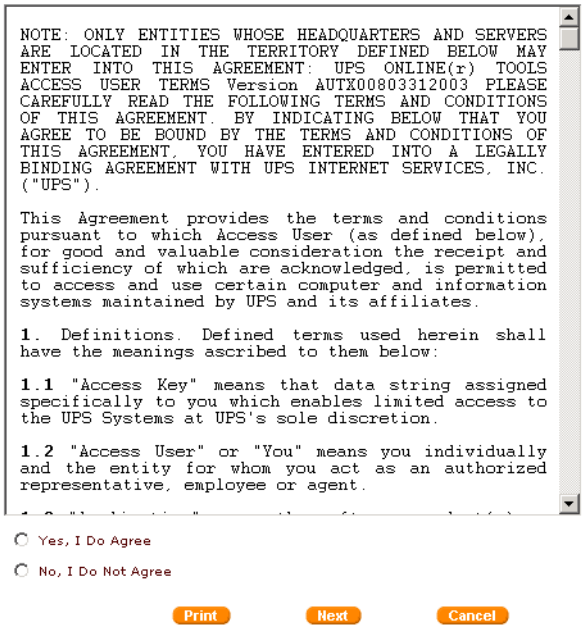
What data is being gathered

- Smart devices
  - What data is transmitted?
- Terms of use documents
  - Lengthy
  - Hard to read
  - Difficult to understand
  - Change often

The Wall Street Journal: the examination of 101 popular smartphone apps revealed that:

- 56 apps transmitted the phone's unique device ID to other companies without users' awareness or consent
- 47 apps transmitted the phone's location in some way
- 5 sent age, gender and other personal details to outsiders

FoxNews: 7.500 online shoppers sold their souls to an online game company on April fool's day 2010
Purple: 22.000 users agreed to 1.000 hours of community service (including cleaning animal waste and relieving sewer blockages) in exchange for free wifi
NCC: reading T&C for an average Norwegian would take 32 hours (250.000 words)



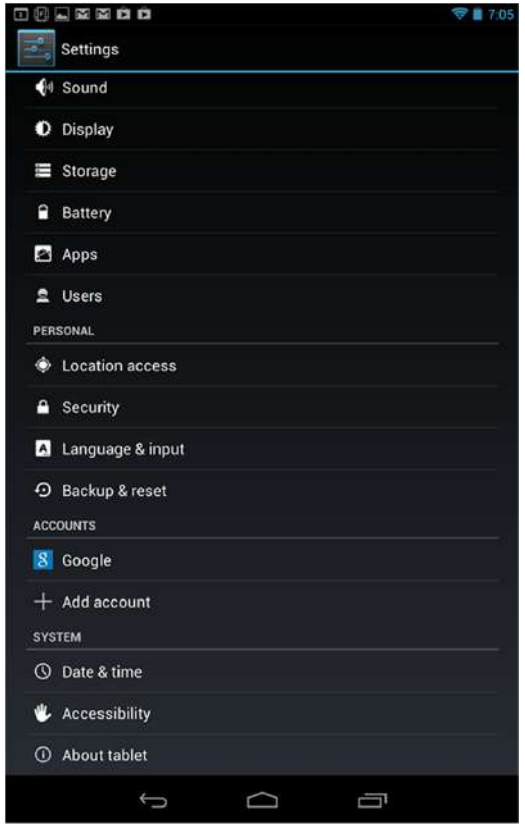
What can be done with data

- The “big data” era
  - Data from everything combined
  - Hidden correlations among seeming unrelated data
  - Impressive deductive capacity

Bloomberg: smart meters can profile homes and habits, including what you watch on TV (via device profiling of energy consumption)
Personality identification: online services can analyze your personality based on authored text
ApplyMagicSauce: can tell your personality from facebook/twitter posts
Vote manipulation: unproven allegations that Cambridge Analytica and other big data companies used targeted micro-advertising and personalized emotional triggers to help in the success of the Brexit and Trump campaigns

Use of mitigating solutions

- Default behaviour of digital objects is not necessarily privacy-respecting
- Technical solutions are built by technical people
  - Not necessarily easy to use or configure
  - Use technical jargon
  - Might lead to loss of functionality
- The average user is not competent enough to do this



People do care

- Three levels of non-awareness:
  - What data is being gathered
    - Unclear, evolving terms of use, technical incompetency
  - What can be done with this data
    - Unaware of technological advances in data analytics
  - Unaware of related mitigating technical solutions
    - Hard to use, hard to find, specific, use technical jargon
- Behavioural difference between digital and real life
- Thus:
  - **People generally do care about their privacy**
  - **The perceived carelessness is due to ignorance**

Bloomberg: according to a survey of nearly 2.300 adults, more than half Americans owning smartphone/tablet have uninstalled an app or declined to download one because of worries about sharing personal information

#2: Can't fight the big companies?

- Many companies earn money out of users' data
  - A novel, lucrative and very successful business model
- If you can't beat them, join them …
  - … or make them join you …
- Use the forces of the market
  - Awareness can lead the public to more privacy-respecting products or services
    - Or maybe not, in which case we are fighting the wrong cause
  - Respect for privacy can be a competitive advantage
    - Not doing that will hurt the market in the long-run
    - Similar case: labour law

#3: Inappropriate legal system?

- Legal frameworks exist
  - GDPR
    - Some companies don't care
  - Central public authority to monitor digital products (Yann Bonnet – French Digital Council)
- Top-down versus bottom-up
  - Policy making is a few steps behind technology
  - For a lasting effect, people's attitude has to change
    - Through awareness

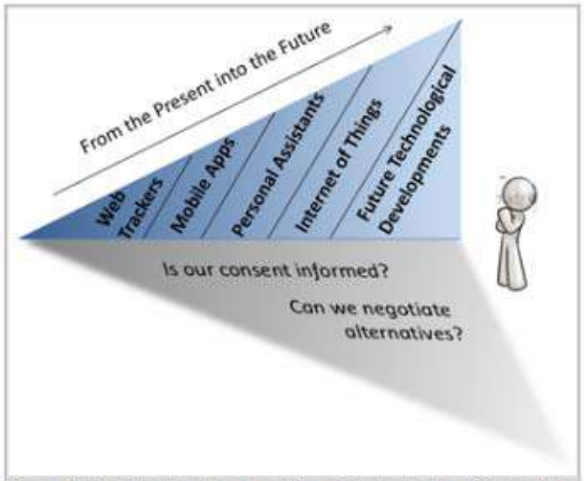


Figure 1. Web trackers (e.g., cookies, widgets, trackers for analytics and advertising), apps for mobile devices, personal assistant software (e.g., Siri, Cortana, Google Now, Alexa), IoT devices (e.g., smart TVs, smart watches, smart cardiac devices etc.), and future technological developments (e.g., social robotics) monitor, process and often share an increasing degree of personal and sensitive data, raising concerns about their legitimate use.

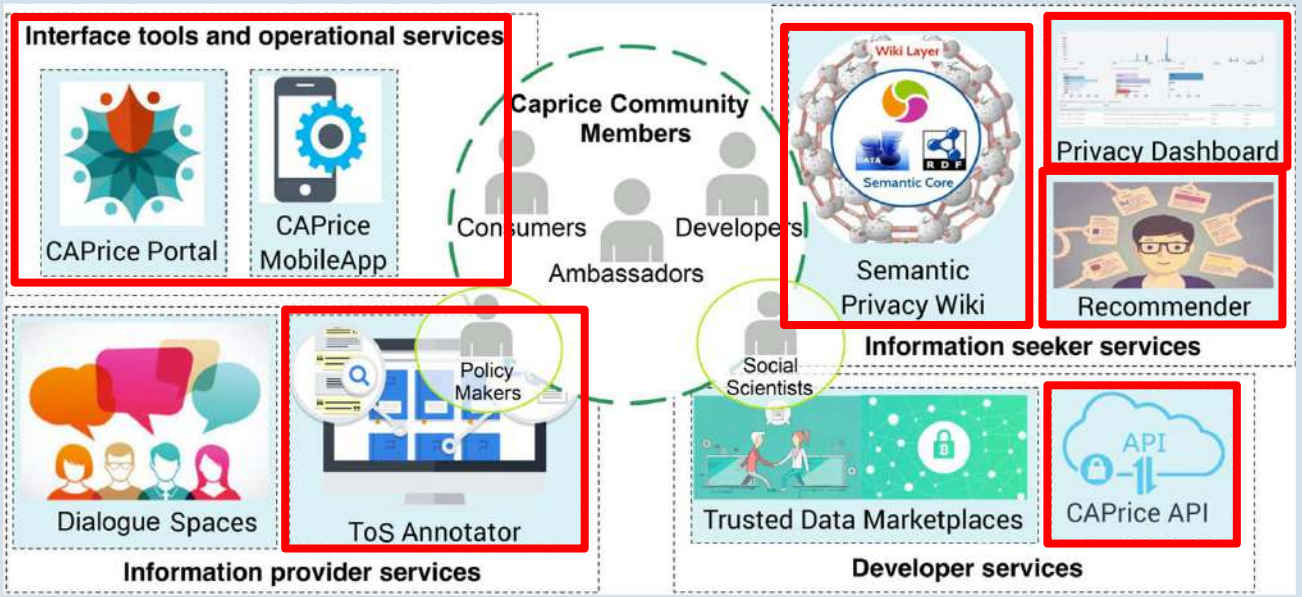
CAPrice: motivation and plan

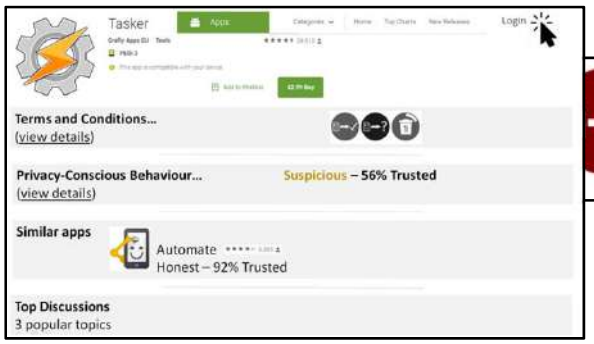
- Non-awareness leads to carelessness
  - Awareness can prevent or mitigate privacy threats
- The plan for our **socio-technical** solution:
  - Build a community of privacy-sensitive individuals
    - Social networking, web site, direct contacts, …
    - Video: https://www.caprice-community.net/idea/
  - ICT tools to support collaboration and awareness
  - Awareness will lead to change in consuming habits
  - The market will adapt, leading to new business models
  - Legislators and policy-makers will follow
- We seek funding (EC, H2020)

CAPrice
collective awareness platform for privacy concerns and expectations

The CAPrice solution

CAPrice
collective awareness platform for privacy concerns and expectations





Giorgos Flouris, July 2017

15

Digital Bias

The problem, and why it is difficult

Joint work with:
Panagiotis Papadakos, Irini Fundulaki

Giorgos Flouris, July 2017

Information needs and bias

- Most information needs are satisfied online
  - Search engines, social networks, e-shops, aggregators, portals
- Many real-life decisions are guided by online searches
  - Hotels, restaurants, investments, health, news, …
  - How trustworthy are they?
    - Correct information?
    - Full spectrum of choices?
    - In the "proper" order?
    - Are the errors systematic (biased)?
- A major AI problem for the coming years

Proving digital bias

- Many accusations
  - Google
    - Biased against Donald Trump?
    - Sexist and racist for some queries (e.g., "doctor", "cop")?
    - Prejudiced about the holocaust?
    - Women less likely to be shown high-wage jobs?
    - Deliberate manipulation of results' ranking (proved in court!)
  - Similar for Facebook, Airbnb, LinkedIn, …
- Can we prove/disprove these accusations?
  - Not easily
  - Legal frameworks exist
  - But hard to prove it from a technical perspective

Why this is a difficult problem

- No ground truth
- Non-analytical and/or secret algorithms
  - Search is typically based on machine learning algorithms
  - Trade secrets: complex filtering and optimization methods
- Training bias (caused by training set)
  - Voice recognition
  - Tay: a sexist, racist chat-bot for twitter
- Dataset bias (dataset distribution)
- User bias (caused by users)
  - Previous clicks by users on similar searches
  - Personalization based on user profile
  - Deliberate manipulation of the PageRank metric

19

Conclusion

Concluding remarks

- Data science and technology in general
  - Tremendous opportunities
  - Great ethical and legal challenges
    - E.g., privacy and digital bias
    - CAPrice: https://www.caprice-community.net/
- Appropriate legislation is necessary, but:
  - Cannot cope with the pace of technology
  - Too rigid for the complex digital world
- Education and awareness is key
  - Ethical and legal issues should be part of the university curricula for technological sciences