Privacy and Security in Online Social Media

Course on NPTEL

NOC21-CS28

Week 4.1

Ponnurangam Kumaraguru (“PK”)
Full Professor
ACM Distinguished Speaker
fb/ponnurangam.kumaraguru, @ponguru

Topics that we will cover

- Overview of OSM
- Linux / Python / Twitter API / Mongo DB / MySQL [Hands-on]
- Trust & Credibility
- **Privacy**
- Social Network Analysis, NLTK [Hands-on]
- e-crime
- Plotly / Highcharts / Geo-location analysis [Hands-on]
- Policing
- Identity resolution
- What next – Deep learning, machine learning, NLP, Image analysis

Westin's 3 categories

- Fundamentalists, 25%
- Pragmatists, 60%
- Unconcerned, 15%

Internet & Social Media

What do you feel about privacy of your personal information on your OSN?

	Q42, N = 6,855
It is not a concern at all	19.30
Since I have specified my privacy settings, my data is secure from a privacy breach	42.13
Even though, I have specified my privacy settings, I am concerned about privacy of my data	23.84
It is a concern, but I still share personal information	8.02
It is a concern; hence I do not share personal data on OSN	6.71

Internet & Social Media

If you receive a friendship request on your most frequently used OSN, which of the following people will you add as friends?

	Q43, N = 6,929
Person of opposite gender	27.39
People from my hometown	19.51
Person with nice profile picture	10.12
Strangers (people you do not know)	4.99
Somebody, whom you do not know or recognize but have mutual / common friends with	8.31
Anyone	2.99

http://precog.iiitd.edu.in/research/privacyindia/

Hard to define

"Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all."

Robert C. Post, *Three Concepts of Privacy*, 89 Geo. L.J. 2087 (2001).

Control over information

"Privacy is the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."

"…each individual is continually engaged in a personal adjustment process in which he balances the desire for privacy with the desire for disclosure and communication…."

Alan Westin, *Privacy and Freedom,* 1967

Forms of Privacy

- Information
  - Internet
- Communication
  - Telephone
- Territorial
  - Living space
- Bodily
  - Self

Background

- In 2000, 100 billion photos were shot worldwide
- In 2010, 2.5 billion photos per month were uploaded by Facebook users only
- In 2015, 1.8 billion photos uploaded everyday on Facebook, Instagram, Flickr, Snapchat, and WhatsApp
- Facebook, Microsoft, Google, Apple have acquired / licensed products that do Face recognition

Many things are colluding

- Increasing public self-disclosures through online social networks
  - Photos
- Improving accuracy in Face recognition
- Cloud, ubiquitous computing
- Re-identification techniques are getting better

Question

- Can one combine publicly available online social network data with off-the-shelf face recognition technology for
  - Individual re-identification
  - Finding potentially, sensitive information

Goal is to

- Use un-identified source {Match.com, photos from Flickr, CCTVs, etc.} + identified sources {Facebook, Linkedin, Govt. websites, etc.}
- To get some sensitive information of the individual {gender orientation, SSN, Aadhaar #, etc.}

Latanya Sweeney

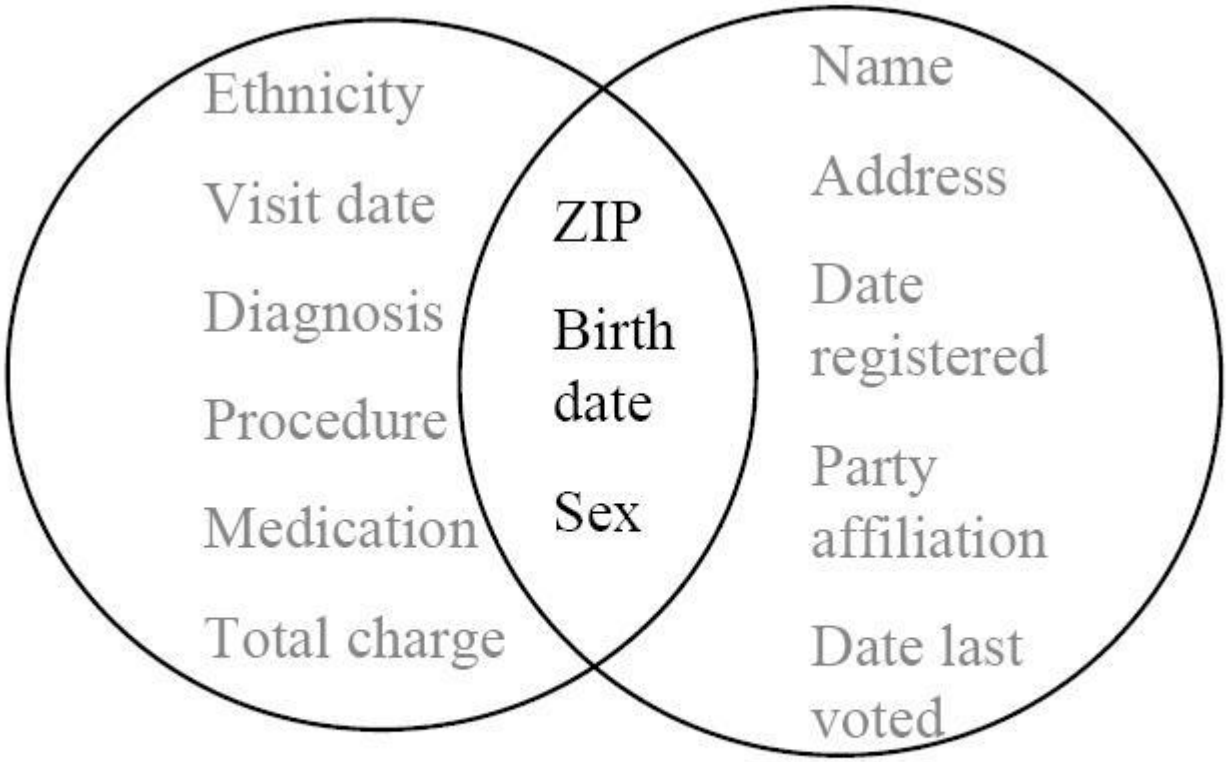


Figure 1 Linking to re-identify data

Experiment 1

- Online – Online
- Mined publicly available images from FB to re-identify profiles on one of the most popular dating sites in the US
- Used http://www.pittpatt.com/ for face recognizing
  - Pittpatt acquired by Google
  - Face detection
  - Face recognition
- Use Tensorflow now

Experiment 1: Data

- Identified
- Downloaded FB profiles from one city in USA
- Profiles: 277,978
- Images: 274,540
- Faces detected: 110,984

Experiment 1: Data

- Un-Identified
- Downloaded profiles of one of the popular dating websites
- Pseudonyms to protect their identities
- Photos can be used to identify
- Same city was used to search
- Profiles: 5,818
- Faces detected: 4,959

Experiment 1: Approach

- Unidentified {Dating site photos} + Identified {FB photos} → Re-identified individual
- More than 500 million pairs compared
- Used only the best matching pair for each dating site picture
- PittPatt produces score of -1.5 to 20
- Crowd sourced to Mturkers for validating PittPatt
- Likert scale, 1 – 5
- At least 5 Turkers for each pair

Experiment 1: Results

- Highly likely matches: 6.3%
- Highly likely + Likely matches: 10.5%
- 1 on 10 from the dating site can be identified

Reactions?

- What can you do better if you were the attacker?

Experiment 2

- Offline to online
- Pictures from FB college network to identify student strolling in campus

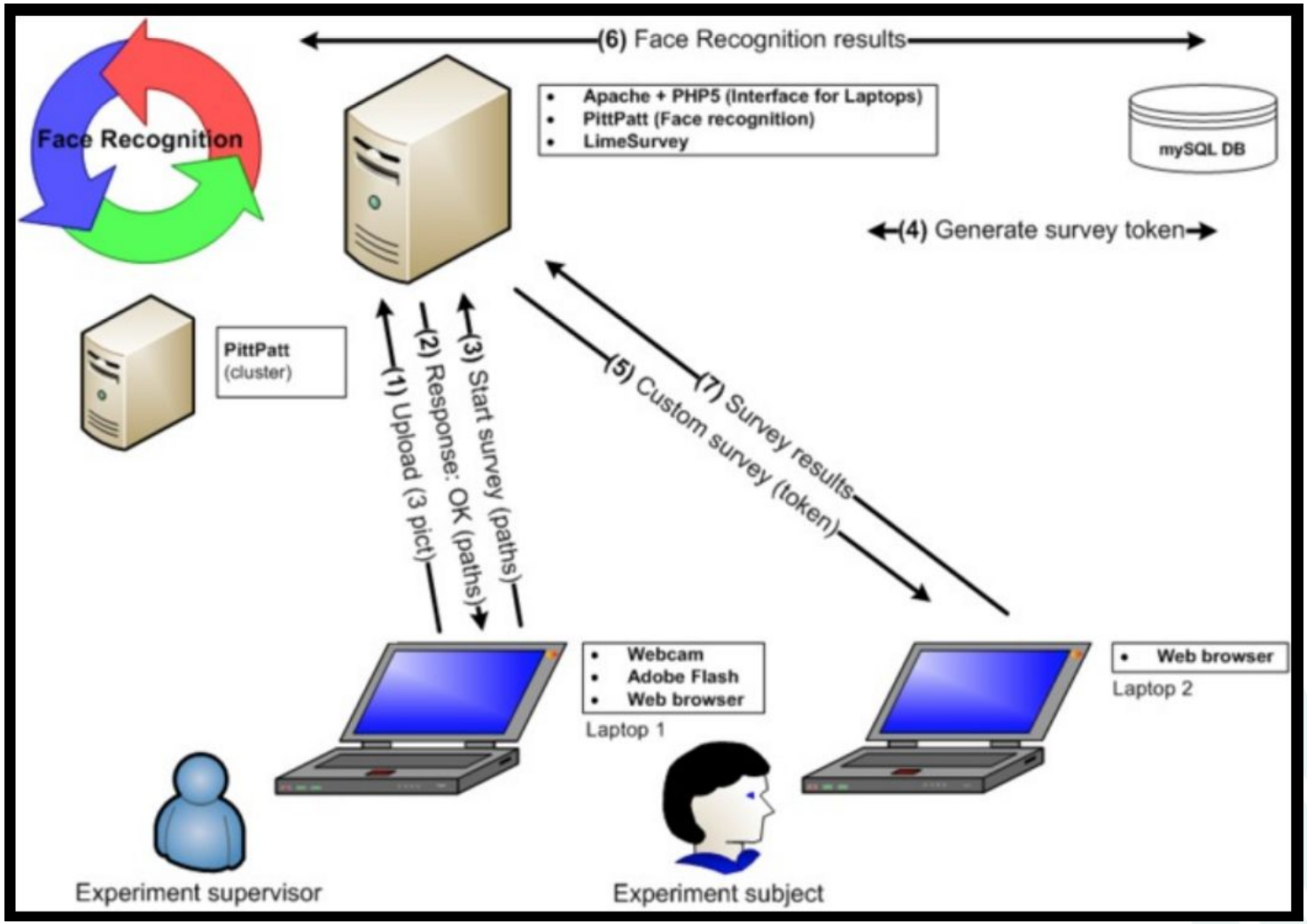
Experiment 2: Data

- Webcam to take 3 pics per participant
- Collected over 2 days
- Facebook data for the university
  - Profiles: 25,051
  - Images: 26,262
  - Faces detected: 114,745

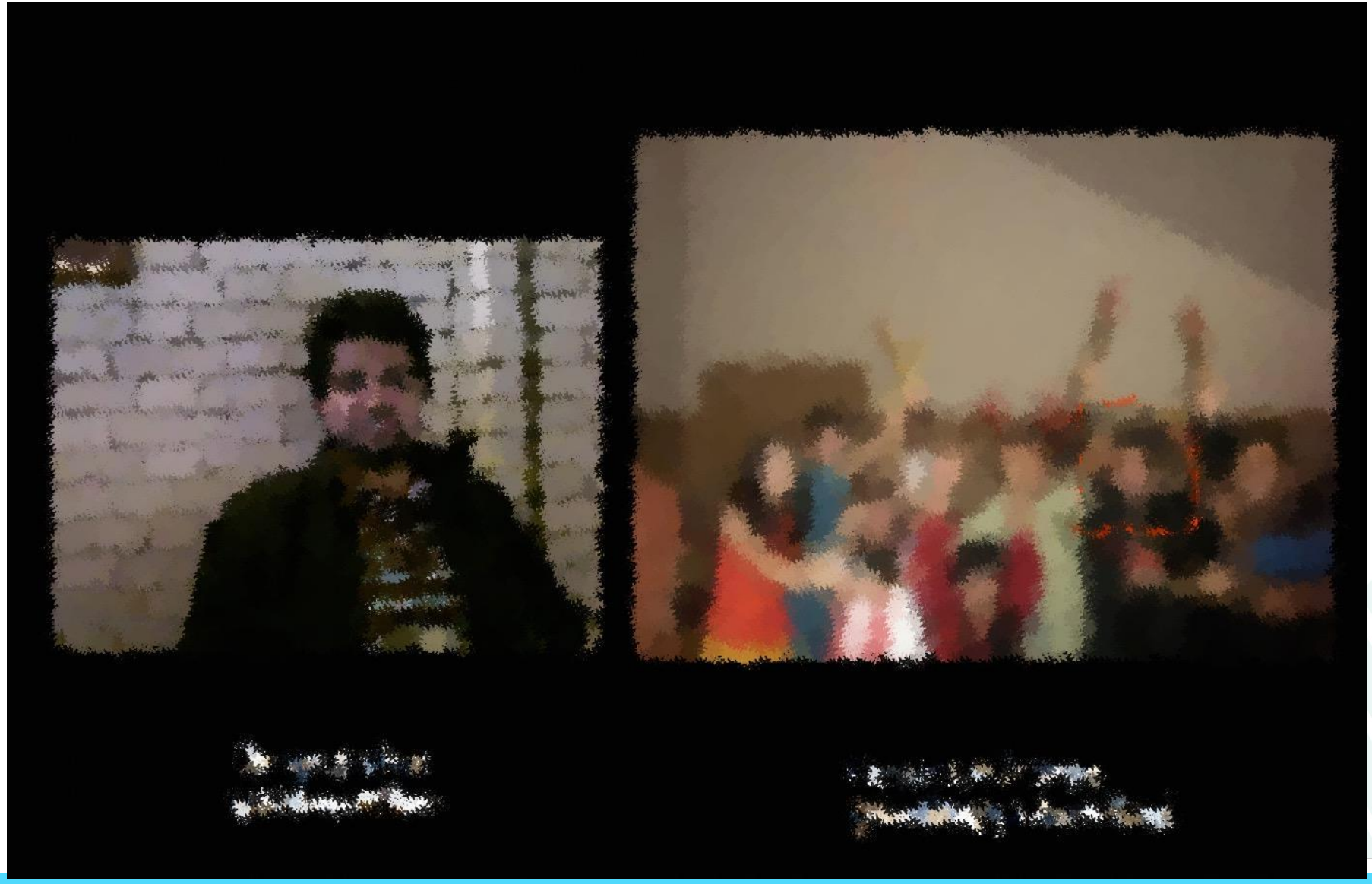
Experiment 2: Process

- Pictures taken of individuals walking in campus
- Asked to fill online survey
- Pictures matched from cloud while they are filling survey
- Last page of the survey with options of their pictures
- Asked to select the pics which matched closely, produced by the recognizer

Experiment 2: Process



Experiment 2: Process



Experiment 2: Results

- 98 participants
  - All students and had FB accounts
- 38.18% of participants were matched with correct FB profile
  - Including a participant who mentioned that he did not have a picture on FB
  - Average computation less than 3 seconds

Experiment 3

- Predicted SSN from public data
- Faces / FB data + Public data ⟶ SSN
- 27% of subjects' first 5 SSN digits identified with four attempts – starting from their faces
- Predicted sensitive information like SSN

What can you think of doing in India?

- Aadhaar number?
- Other details?

References

- https://www.blackhat.com/docs/webcast/acquisti-face-BH-Webinar-2012-out.pdf
- http://www.heinz.cmu.edu/~acquisti/papers/privacy-facebook-gross-acquisti.pdf

Thank you

pk@iiitd.ac.in
precog.iiitd.edu.in
fb/ponnurangam.kumaraguru