Social Footprint

Using Profile Attributes to Determine Vulnerability in Online Social Networks

Group 13

Hamza Karachiwala

Kedar Amrolkar

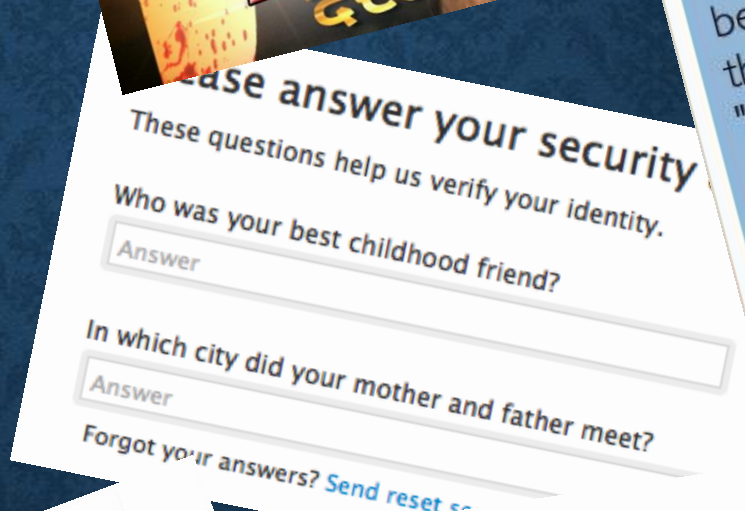
Mickey Vellukunnel

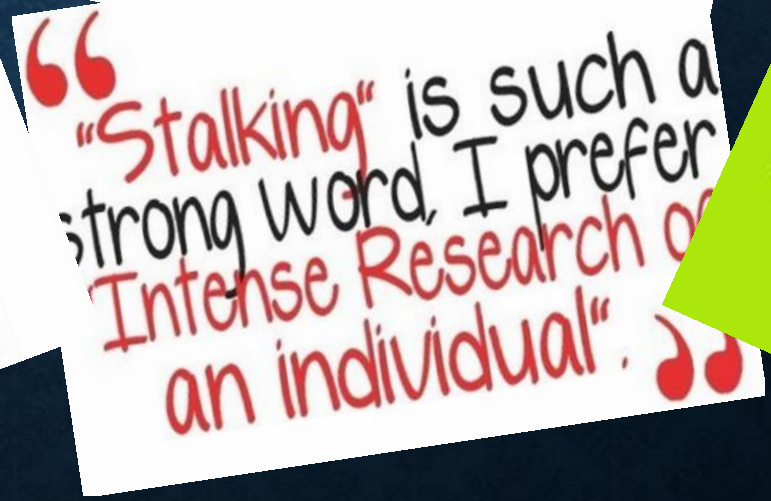
SOCIAL FOOTPRINT
PROFILE INFORMATION AGGREGATED FROM MULTIPLE OSN ACCOUNTS

- How much aggregated information do we disclose online through multiple Online Social Networking accounts?
- How safe is it to disclose that information? Can we measure the risk associated with it?
- Can we identify a way to reduce disclosed information but keep visibility high?

MOTIVATION

- More comprehensive profile out there than we expect.
- Identity thefts
- Password Recovery
- Stalking / Harassment
- Sex crimes





PROPOSED SOLUTION

- Use OSN APIs for developers to fetch profile information of a user.
- Aggregate data from all APIs to create the social footprint.
- Have weights associated with relevant/crucial available public attributes.
- Calculate total weight of the footprint and compare with a threshold.
- Identify attributes that could be left out to reduce weight (associated with risk) with less effect on visibility.

SOCIAL NETWORKS OUT THERE

- Facebook - Most personal data disclosed due to popularity.
- LinkedIn - Professional data usually not available on Facebook.
- Quora - Also popular and categorized.
- Instagram - Images.
- YouTube - Videos.
- Google+ - Does anybody use it.
- Twitter - Followers.
- Pinterest - Interests.








WEIGHTS FOR PROFILE ATTRIBUTES

- Weights based on risk associated with attributes.
- The riskier an attribute the more weight it will have.
- These could be risks of
  - Password recovery
  - Identity theft & Profiling
  - PII - Personal Identification Information (using Cross Site Linking)

How we assign weights to attributes

- Look at previous research to see how different fields are used in different sites or across sites.
- Use the balance beam approach to assign weights based on these percentages/records.
- Use the weighted averages approach to decide how much risk a particular sub-category of crime warrants.

Example

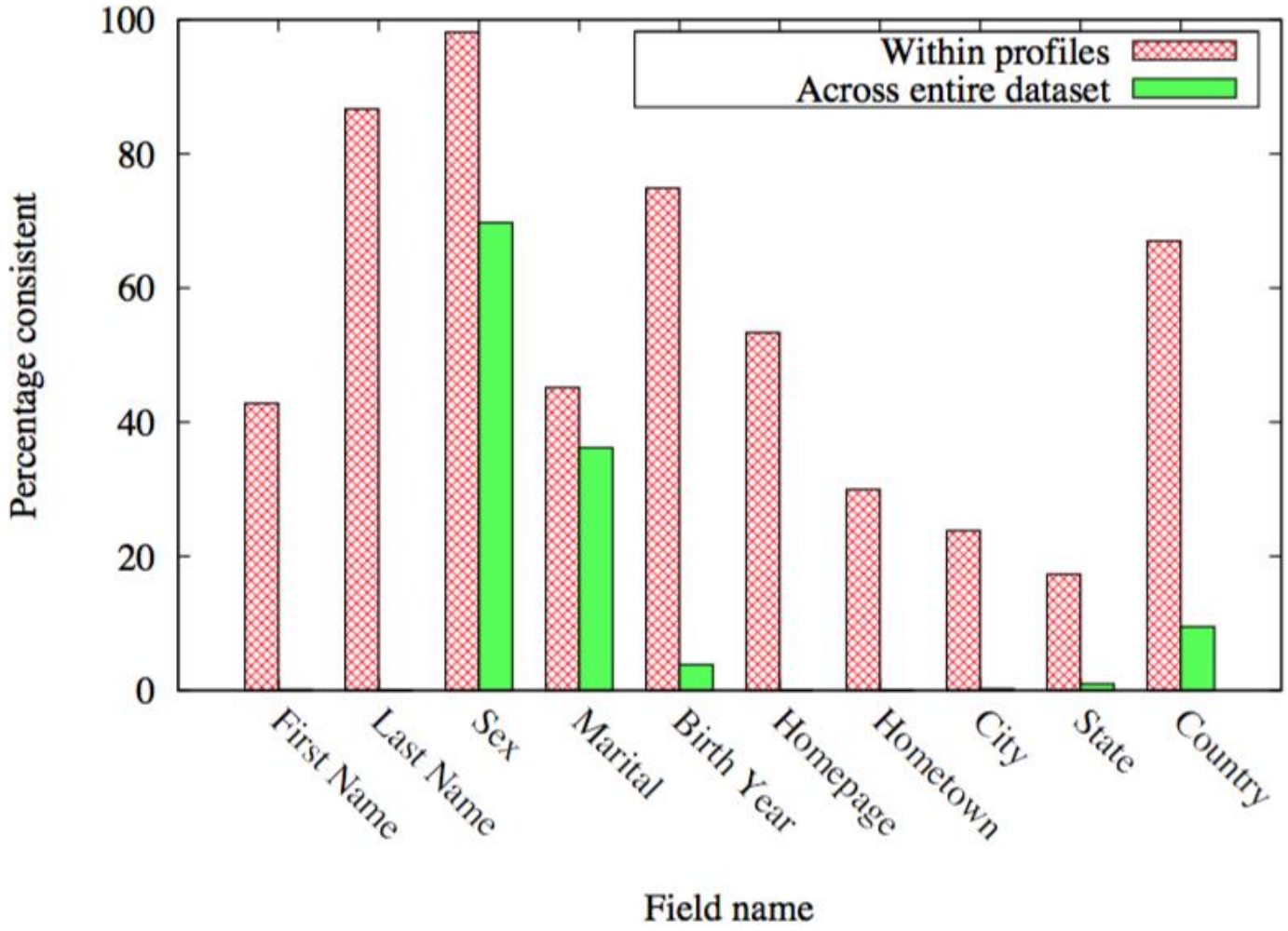


Fig. 7. Consistency of fields within a profile and across the entire dataset.

Example of weights calculated...

Field	Weight associated
Email ID	0.35
Name	0.21
Profile picture	0.17
Date of birth	0.12
Zipcode	0.07
Hometown	0.06
Age	0.01
Location	0.01

WHY PASSWORD RECOVERY

- Common misconception; I've got a strong password, my profile is safe.
- What about password recovery questions?
  - Sarah Palin's "hacked" email account.
  - 80% people have admitted to using recoveries
  - Not as safe as other recovery techniques
- We generally choose easy to remember questions.

PASSWORD RECOVERY QUESTIONS

- City of Birth?
- First School?
- Mothers maiden name?
- First kiss from?
- Favourite restaurant, actor, artist?
- Wedding Reception Venue?
- Father's middle name?
- Pets Name?
- First Car?
- First manager at job?
- Vehicle registration number?
- Wedding Reception Venue?
- Parents' birth year?

TECH STACK

- PHP - Use the APIs provided by OSNs to fetch user data
  - Facebook
  - LinkedIn
- MySQL - Store the footprint and weights of attributes
  - Store footprint weights in the backend - statistical information
  - Calculate the weight and compare with threshold on server
- Apache now, host later
  - We could move from localhost to the cloud if time permits

STALKING THE HAMZA

- Hidden from FB, visible on LinkedIn
  - Email
  - Location

Current User: hskarachiwala Log Out

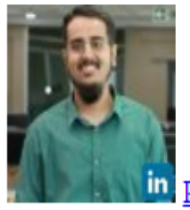
FACEBOOK

Profile URL

Name: Hamza Karachiwala
Gender: male
Email Id: HIDDEN
Hometown: HIDDEN
Current location: HIDDEN
DOB: HIDDEN
Locale: en_GB
Languages: HIDDEN
Checkins:

PLACE	DATE	TIME
Gateway at Glades	2015/09/24	02:09:13
Gateway at Glades	2015/09/11	05:09:22
Southwest Recreation Center	2015/08/21	21:08:08
Zs Associates, Pentagon 5, Magarpatta Cybercity , Pune	2015/07/03	14:07:41
Pasteur	2015/05/08	18:05:56
NH37 Dhaba	2015/03/23	17:03:04
The Brooklyn Shuffle	2015/03/08	15:03:14
Seasons Mall	2015/03/07	10:03:01
Zs Associates, Pentagon 5, Magarpatta Cybercity , Pune	2015/02/24	09:02:26

LINKEDIN



Profile URL

First Name: Hamza
LastName: Karachiwala
headline: Graduate Computer Science Student at the University of Florida
UserId: HlClvbmYmG
Location: Gainesville, Florida Area
Industry: Computer Software
Summary: Graduate student in the Department of Computer Science at the University of Florida. Expecting to graduate in December 2016. Currently have two years' industry experience in web application development.
EmailAddress: hskarachiwala@gmail.com

What has changed in our scope? Challenges?

- Deriving weights from crime records.
- Giving risk assessments for “friend” visible information.
- Informing the user of every type of risk associated with a particular field.

ROADMAP

- Use OSN APIs for developers to fetch data for a user - In Progress
- Aggregate data from all APIs to create the total footprint - In Progress
- Have weights associated with all attributes - In Progress
- Calculate total weight of the footprint and compare with a threshold - Pending
- Identify attributes that could be left out to reduce weight with less effect on visibility- Pending

QUESTIONS AND SUGGESTIONS?