Sharing Multimedia on the Internet and the Impact for Online Privacy

Dr. Gerald Friedland

Director Audio and Multimedia Group
International Computer Science Institute

Berkeley, CA

friedland@icsi.berkeley.edu

Question

Question

On average, how often are you posting images and videos on the Internet (e.g. Facebook, Flickr, Craigslist) ?

Question

On average, how often are you posting images and videos on the Internet (e.g. Facebook, Flickr, Craigslist) ?

a) Never

Question

On average, how often are you posting images and videos on the Internet (e.g. Facebook, Flickr, Craigslist) ?

    a) Never

    b) About once a month or less

Question

On average, how often are you posting images and videos on the Internet (e.g. Facebook, Flickr, Craigslist) ?

    a) Never

    b) About once a month or less

    c) About once a week

Question

On average, how often are you posting images and videos on the Internet (e.g. Facebook, Flickr, Craigslist) ?

　　a) Never

　　b) About once a month or less

　　c) About once a week

　　d) About once a day

Question

On average, how often are you posting images and videos on the Internet (e.g. Facebook, Flickr, Craigslist) ?

    a) Never

    b) About once a month or less

    c) About once a week

    d) About once a day

    e) More than once a day

A Popular Introduction to the Problem

Question

Question

How would you judge the issue raised by Colbert?

Question

How would you judge the issue raised by Colbert?

  a) It's a comedy. I don't worry about any of this.

Question

How would you judge the issue raised by Colbert?

    a) It's a comedy. I don't worry about any of this.

    b) There is some truth to it but its mostly exaggarated.

Question

How would you judge the issue raised by Colbert?

a) It's a comedy. I don't worry about any of this.

b) There is some truth to it but its mostly exaggarated.

c) It's a comedy depection of the reality but most of the stuff is becoming an issue.

Question

How would you judge the issue raised by Colbert?

a) It's a comedy. I don't worry about any of this.

b) There is some truth to it but its mostly exaggarated.

c) It's a comedy depection of the reality but most of the stuff is becoming an issue.

d) He only touched a small part of the problem. The actual issues are even more serious.

4

Our Observations

Our Observations

- Many Internet sites and mobile apps encourage sharing of data too easily and users follow.

Our Observations

- Many Internet sites and mobile apps encourage sharing of data too easily and users follow.

- Users **and** even engineers often unaware of (hidden) search and retrieval possibilities of shared data.

Our Observations

- Many Internet sites and mobile apps encourage sharing of data too easily and users follow.

- Users **and** even engineers often unaware of (hidden) search and retrieval possibilities of shared data.

- Local privacy protection ineffective against inference across web-sites.

Social Cause

Social Cause

●People want to post on the Internet and like a highly-personalized web experience.

Social Cause

● People want to post on the Internet and like a highly-personalized web experience.

● Industry is improving search and retrieval techniques so that people can find the posts.

Social Cause

- People want to post on the Internet and like a highly-personalized web experience.
- Industry is improving search and retrieval techniques so that people can find the posts.
- Governments improve search and retrieval to do forensics and intelligence gathering

Let's focus

Let's focus

- The previous described issues are a problem with any type of public or semi-public posts and are not specific to a certain type of information, e.g. text, image, or video.

Let's focus

- The previous described issues are a problem with any type of public or semi-public posts and are not specific to a certain type of information, e.g. text, image, or video.

- However, let's focus on multimedia data: images, audio, video.

Multimedia in the Internet is Growing

Multimedia in the Internet is Growing

- YouTube claims ~~65k~~ 100k video uploads per day, or ~~48h~~ 72h per minute.

Multimedia in the Internet is Growing

- YouTube claims ~~65k~~ 100k video uploads per day, or ~~48h~~ 72h per minute.

- Flickr claims 1M images uploads per day

8

Multimedia in the Internet is Growing

- YouTube claims ~~65k~~ 100k video uploads per day, or ~~48h~~ 72h per minute.

- Flickr claims 1M images uploads per day

- Twitter: up to 120M messages per day

Resulting Problem

Resulting Problem

- More multimedia data = Higher demand for retrieval and organization tools.

Resulting Problem

- More multimedia data = Higher demand for retrieval and organization tools.
- But multimedia retrieval is hard

Resulting Problem

- More multimedia data = Higher demand for retrieval and organization tools.

- But multimedia retrieval is hard
  ➡Researchers work on making retrieval better.

Resulting Problem

- More multimedia data = Higher demand for retrieval and organization tools.

- But multimedia retrieval is hard
  - ➡Researchers work on making retrieval better.
  - ➡Industry develops workarounds to make retrieval easier right away.

Hypothesis

Hypothesis

- Retrieval is already good enough to cause major issues for privacy that are not easy to solve.

Hypothesis

- Retrieval is already good enough to cause major issues for privacy that are not easy to solve.

- Let's take a look at some retrieval approaches:

Hypothesis

- Retrieval is already good enough to cause major issues for privacy that are not easy to solve.

- Let's take a look at some retrieval approaches:
  - Geotagging

Hypothesis

- Retrieval is already good enough to cause major issues for privacy that are not easy to solve.

- Let's take a look at some retrieval approaches:
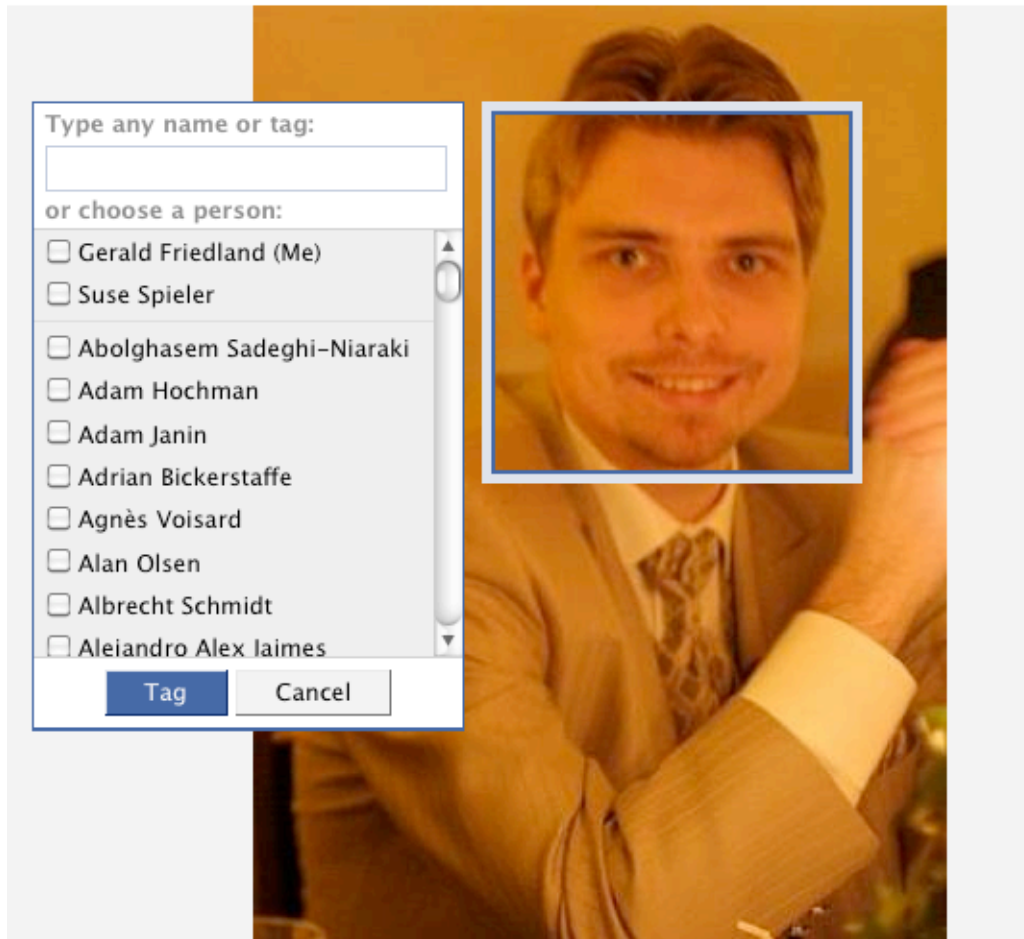 - Geotagging
  - Multimodal Location Estimation

Hypothesis

- Retrieval is already good enough to cause major issues for privacy that are not easy to solve.

- Let's take a look at some retrieval approaches:
  - Geotagging
  - Multimodal Location Estimation
  - Audio-based user matching

Workaround: Manual Tagging

Workaround: Geotagging

Source: Wikipedia

Geo-Tagging

Allows easier clustering of photo and video series as well as additional services.

Support for Geo-Tags

Social media portals provide APIs to connect geo-tags with metadata, accounts, and web content.

Portal	%	Total
YouTube (estimate)	3.0	3M
Flickr	4.5	180M

Allows easy search, retrieval, and ad placement.

Related Work

"Be careful when using social location sharing services, such as FourSquare."

Related Work

Mayhemic Labs, June 2010: "Are you aware that Tweets are geo-tagged?"

Question

Question

Did you know about geo-tagging and its potential?

Question

Did you know about geo-tagging and its potential?

a) I had never heard about geo-tagging before.

Question

Did you know about geo-tagging and its potential?

> a) I had never heard about geo-tagging before.

> b) I knew about geo-tagging but never thought about what it could be used for.

Question

Did you know about geo-tagging and its potential?

    a) I had never heard about geo-tagging before.

    b) I knew about geo-tagging but never thought about what it could be used for.

    c) I knew about geo-tagging and knew the potential for photo organization and retrieval

Question

Did you know about geo-tagging and its potential?

a) I had never heard about geo-tagging before.

b) I knew about geo-tagging but never thought about what it could be used for.

c) I knew about geo-tagging and knew the potential for photo organization and retrieval

d) I know about geo-tagging, it's use and the privacy risks.

Question

Did you know about geo-tagging and its potential?

a) I had never heard about geo-tagging before.

b) I knew about geo-tagging but never thought about what it could be used for.

c) I knew about geo-tagging and knew the potential for photo organization and retrieval

d) I know about geo-tagging, it's use and the privacy risks.

e) I only heard about privacy risks of geo-tagging but never really thought about what it is good for.

Hypothesis

Hypothesis

Since geo–tagging is a workaround for multimedia retrieval, it allows us to peek into a future where multimedia retrieval works.

Hypothesis

Since geo-tagging is a workaround for multimedia retrieval, it allows us to peek into a future where multimedia retrieval works.

Hypothesis

Since geo-tagging is a workaround for multimedia retrieval, it allows us to peek into a future where multimedia retrieval works.

What if multimedia retrieval actually worked?

Can you do real harm?

G. Friedland and R. Sommer: "Cybercasing the Joint: On the Privacy Implications of Geotagging", Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10), Washington, D.C, August 2010. [19]

Can you do real harm?

- **Cybercasing: Using online (location-based) data and services to enable real-world attacks.**

G. Friedland and R. Sommer: "Cybercasing the Joint: On the Privacy Implications of Geotagging", Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10), Washington, D.C, August 2010. [19]

Can you do real harm?

- **Cybercasing: Using online (location-based) data and services to enable real-world attacks.**

- **Three Case Studies:**

G. Friedland and R. Sommer: "Cybercasing the Joint: On the Privacy Implications of Geotagging", Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10), Washington, D.C, August 2010.

Can you do real harm?

- **Cybercasing: Using online (location-based) data and services to enable real-world attacks.**

- **Three Case Studies:**

G. Friedland and R. Sommer: "Cybercasing the Joint: On the Privacy Implications of Geotagging", Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10), Washington, D.C, August 2010. [19]

Case Study 1: Twitter

Case Study 1: Twitter

- Pictures in Tweets can be geo-located

Case Study 1: Twitter

• Pictures in Tweets can be geo-located

Case Study 1: Twitter

- Pictures in Tweets can be geo-located

- From a technically-savvy celebrity we found:

Case Study 1: Twitter

- Pictures in Tweets can be geo-located

- From a technically-savvy celebrity we found:
  - Home location (several pics)

Case Study 1: Twitter

- Pictures in Tweets can be geo-located

- From a technically-savvy celebrity we found:
  – Home location (several pics)
  – Where the kids go to school

Case Study 1: Twitter

- Pictures in Tweets can be geo-located

- From a technically-savvy celebrity we found:
  - Home location (several pics)
  - Where the kids go to school
  - The place where he/she walks the dog

Case Study 1: Twitter

- Pictures in Tweets can be geo-located

- From a technically-savvy celebrity we found:
  - Home location (several pics)
  - Where the kids go to school
  - The place where he/she walks the dog
  - "Secret" office

Case Study 1: Twitter

- Pictures in Tweets can be geo-located

- From a te[...] [...]rity we found:
  - Home lo[...]
  - Where th[...]
  - The place where he/she walks the dog
  - "Secret" office

Celebs unaware of Geo-Tagging

twitpic

Click here to login or

Working with the very talented Adam Hamilton on creating a new album. My best, Bill

Celebs unaware of Geotagging

EXIF IFD1

- Compression {0x0103} = JPEG compression (6)
- X-Resolution {0x011A} = 4718592/65536 ===> 72
- Y-Resolution {0x011B} = 4718592/65536 ===> 72
- X/Y-Resolution Unit {0x0128} = inch (2)
- Y/Cb/Cr Positioning (Subsampling) {0x0213} = centered / center of pixel array (1)
- Embedded thumbnail image:

EXIF GPS IFD

- GPS Version ID {0x00} = 0x02,0x02,0x00,0x00
- GPS Latitude Reference {0x01} = N
- GPS Latitude {0x02} = 34/1,12/1,3/1 [degrees, minutes, seconds] ===> 34° 12′ 3″ == 34.200833°
- GPS Longitude Reference {0x03} = W
- GPS Longitude {0x04} = ████████████ [degrees, minutes, seconds] ===> ██████████ ′ == ████████

Google Maps shows Address...
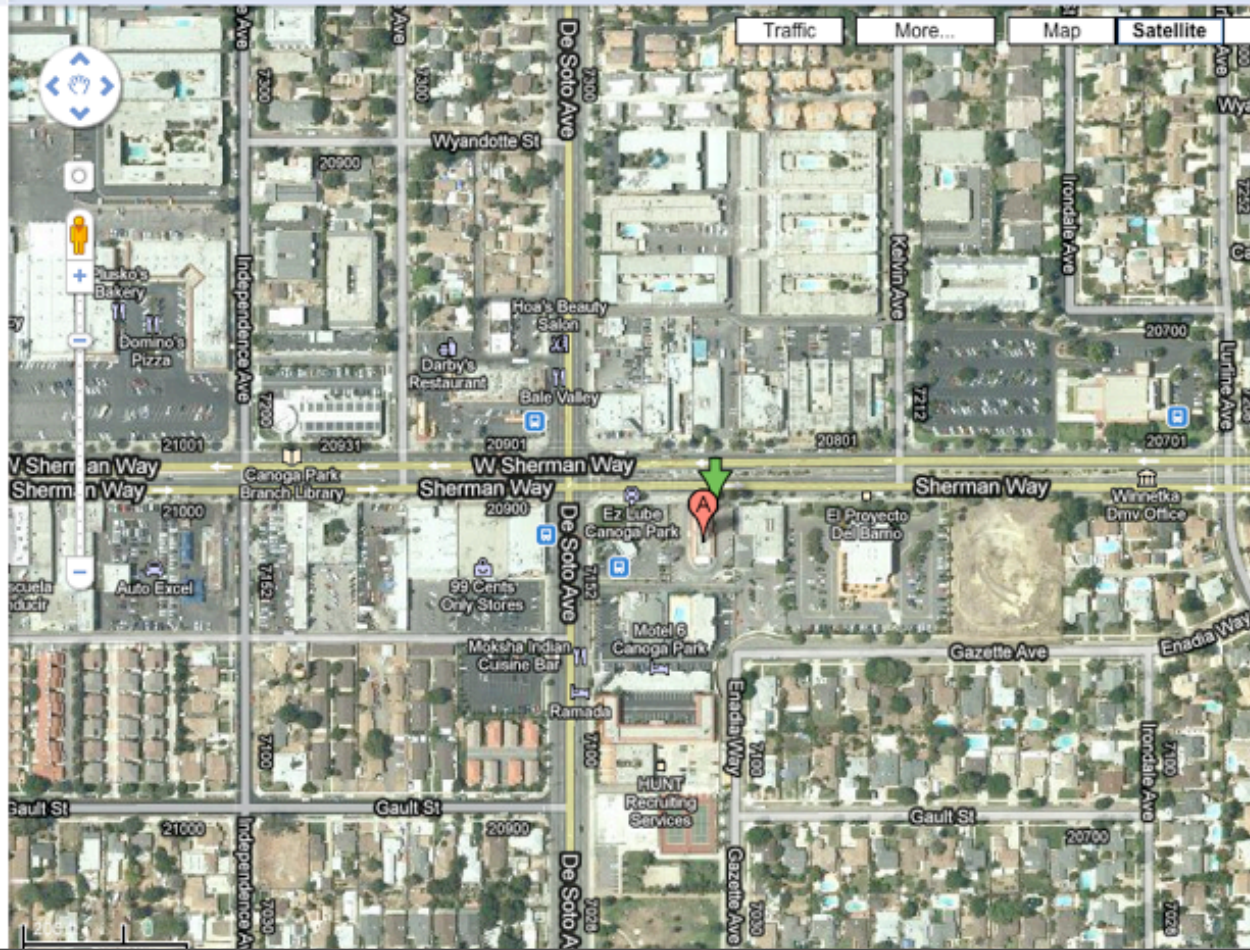
Case Study 2: Craigslist

"For Sale" section of Bay Area <u>Craigslist.com</u>:
4 days: 68729 pictures total,1.3% geo-tagged

#	Model	#	Model
414	iPhone 3G	6	Canon PowerShot SD780
287	iPhone 3GS	3	MB200
98	iPhone	2	LG LOTUS
32	Droid	2	HERO200
26	SGH-T929	2	BlackBerry 9530
20	Nexus One	1	RAPH800
9	SPH-M900	1	N96
9	RDC-i700	1	DMC-ZS7
6	T-Mobile G1	1	BlackBerry 9630

People are Unaware of Geo-Tagging

People are Unaware of Geo-Tagging

- Many ads with geo-location otherwise anonymized

People are Unaware of Geo-Tagging

- Many ads with geo-location otherwise anonymized
- Sometimes selling high-valued goods, e.g. cars, diamonds

People are Unaware of Geo-Tagging

- Many ads with geo-location otherwise anonymized
- Sometimes selling high-valued goods, e.g. cars, diamonds
- Sometimes "call Sunday after 6pm"
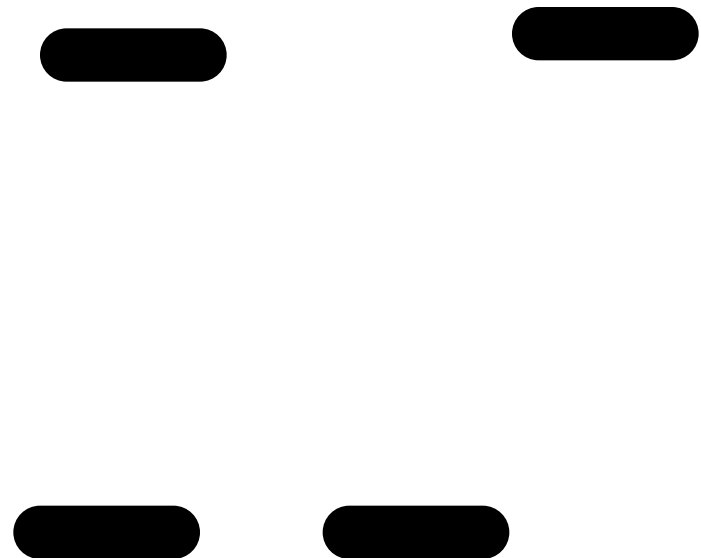
People are Unaware of Geo-Tagging

- Many ads with geo-location otherwise anonymized

- Sometimes selling high-valued goods, e.g. cars, diamonds

- Sometimes "call Sunday after 6pm"

- Multiple photos allow interpolation of coordinates for higher accuracy

Craigslist: Real Example

Craigslist: Real Example

Case Study 3: YouTube

Case Study 3: YouTube

Recall:

Case Study 3: YouTube

Recall:

● Once data is published, the Internet keeps it (in potentially many copies).

Case Study 3: YouTube

Recall:

• Once data is published, the Internet keeps it (in potentially many copies).

• APIs are easy to use and allow quick retrieval of large amounts of data
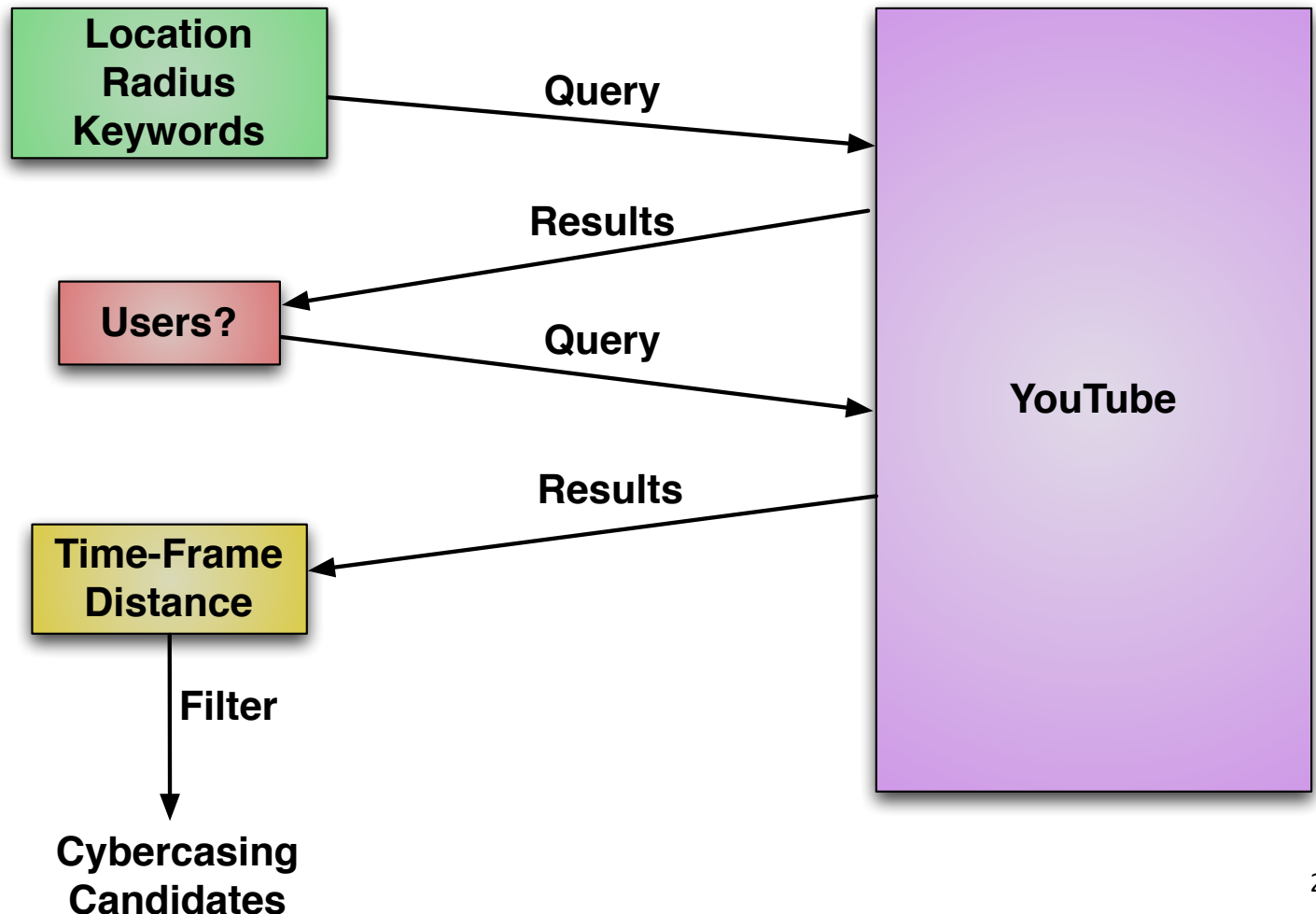
Case Study 3: YouTube

Recall:

• Once data is published, the Internet keeps it (in potentially many copies).

• APIs are easy to use and allow quick retrieval of large amounts of data

Can we find people on vacation in YouTube?

Cybercasing on YouTube

Experiment: Cybercasing using the YouTube API (240 lines in Python)

Cybercasing on YouTube

Input parameters

Cybercasing on YouTube

Input parameters

Location: `37.869885,-122.270539`
Radius: `100km`
Keywords: `kids`
Distance: `1000km`
Time-frame: `this_week`

Cybercasing on YouTube

Output

Cybercasing on YouTube

Output

Initial videos: `1000 (max_res)`

Cybercasing on YouTube

Output

Initial videos: `1000 (max_res)`

➡ User hull: `~`$50k$ `videos`

Cybercasing on YouTube

Output

Initial videos: `1000 (max_res)`

➡ User hull: `~50k videos`

➡ Vacation hits: `106`

Cybercasing on YouTube

Output

Initial videos: `1000 (max_res)`

➡ User hull: `~50k videos`

➡ Vacation hits: `106`

➡ Cybercasing targets: `>12`

Cybercasing on YouTube

First Day of ███████ Vacation
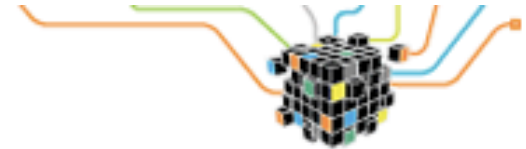
███████████ videos ☒   Subscribe

Ou

In

➡

➡

➡

30

The Threat is Real!

Bits

Business ▪ Innovation ▪ Technology ▪ Society

September 12, 2010, 10:24 AM

Burglars Picked Houses Based on Facebook Updates

By NICK BILTON

📋 **News Feed**   Top News · Most Recent **63**

Share: 📄 Status   💡 Question   🖼 Photo   🔗 Link   📹 Video

> Going to the beach for the weekend! (Someone else will be home though so think again Facebook Bandits!)

🔒 ▾   **Share**

Illustration by Nick Bilton/The New York Times

Question

Question

Do you think geo-tagging should be illegal?

Question

Do you think geo-tagging should be illegal?

a) No, people just have to be more careful. The possibilities still outweigh the risks.

Question

Do you think geo-tagging should be illegal?

a) No, people just have to be more careful. The possibilities still outweigh the risks.

b) Maybe it should be regulated somehow to make sure no harm can be done.

Question

Do you think geo-tagging should be illegal?

    a) No, people just have to be more careful. The possibilities still outweigh the risks.

    b) Maybe it should be regulated somehow to make sure no harm can be done.

    c) Yes, absolutely this information is too dangerous.

But…

But...

Is this really about geo–tags?
(remember: hypothesis)

Ongoing Work:

The Berkeley
Multimodal Location Estimation
Project

http://mmle.icsi.berkeley.edu

Multimodal Location Estimation

We infer location of a Video based on visual, audio, and tags:

●Use geo–tagged data as training data

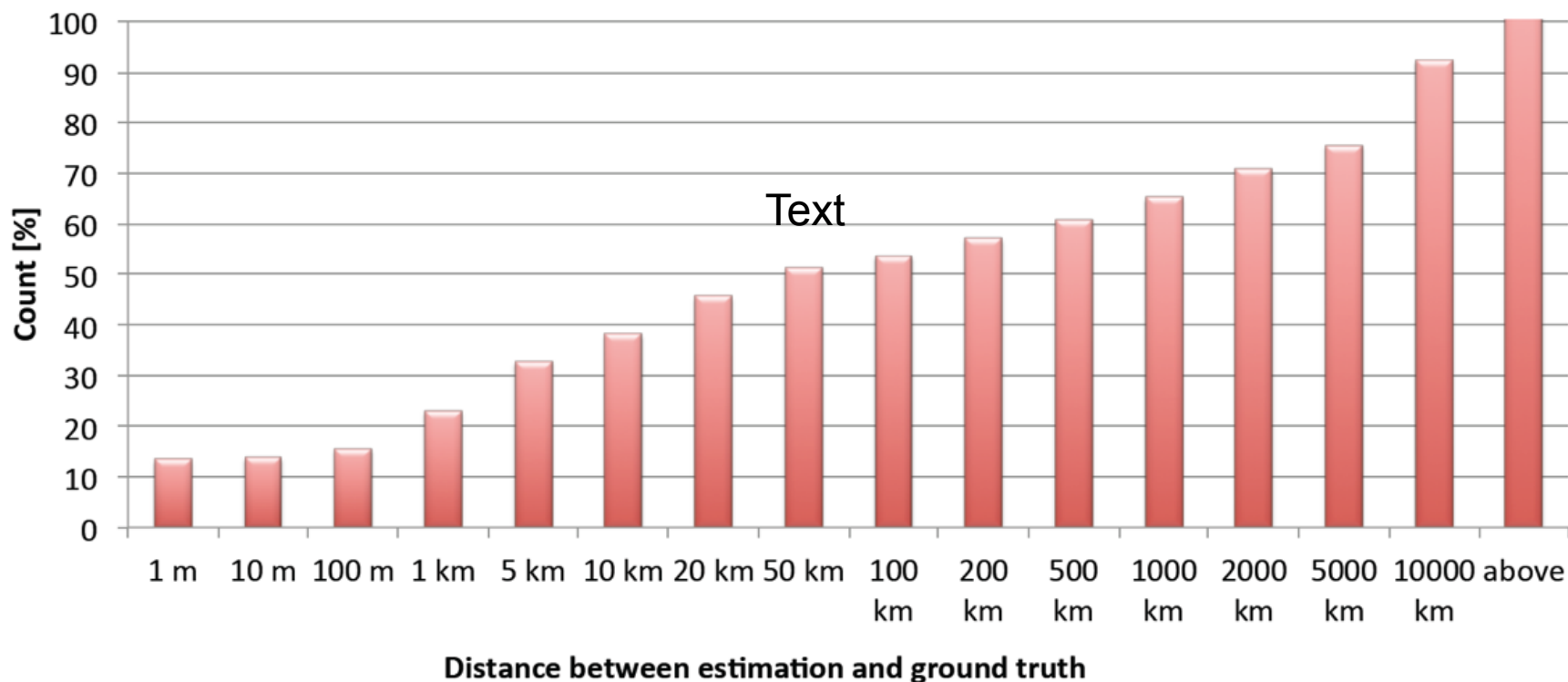●Allows faster search, inference, and intelligence gathering even without GPS.

G. Friedland, O. Vinyals, and T. Darrell: "Multimodal Location Estimation," pp. 1245-1251, ACM Multimedia, Florence, Italy, October 2010.

ICSI/UCB Estimation System at Placing Task 2012 (Cumulative)

J. Choi, G. Friedland, V. Ekambaram, K. Ramchandran: "Multimodal Location Estimation of Consumer Media: Dealing with Sparse Training Data," in Proceedings of IEEE ICME 2012, Melbourne, Australia, July 2012.

YouTube Cybercasing Revisited

	Old Experiment	No Geotags
Initial Videos	1000 (max)	107
User Hull	~50k	~2000
Potential Hits	106	112
Actual Targets	>12	>12

YouTube Cybercasing with Geo-Tags vs Multimodal Location Estimation

Question

Question

Do you think research about geo-location should be abandonend?

Question

Do you think research about geo-location should be abandonend?

   a) No, of course not.

Question

Do you think research about geo-location should be abandonend?

    a) No, of course not.

    b) No, but regulated.

Question

Do you think research about geo-location should be abandonend?

    a) No, of course not.

    b) No, but regulated.

    c) Yes, absolutely.

But…

But...

Is this really only about geo-location?

But...

Is this really only about geo-location?

No, it's about the privacy implications of multimedia retrieval in general.

Example

Example

Idea: Can one link videos across acounts?

Example

Idea: Can one link videos across acounts?

(e.g. YouTube linked to Facebook vs anonymized dating site)

User ID on Flickr videos

User–ID, i–vector system, 6.4 million trials (3287 true positive trials), 1239 train users, 2784 test users, 60–16000 Hz MFCC
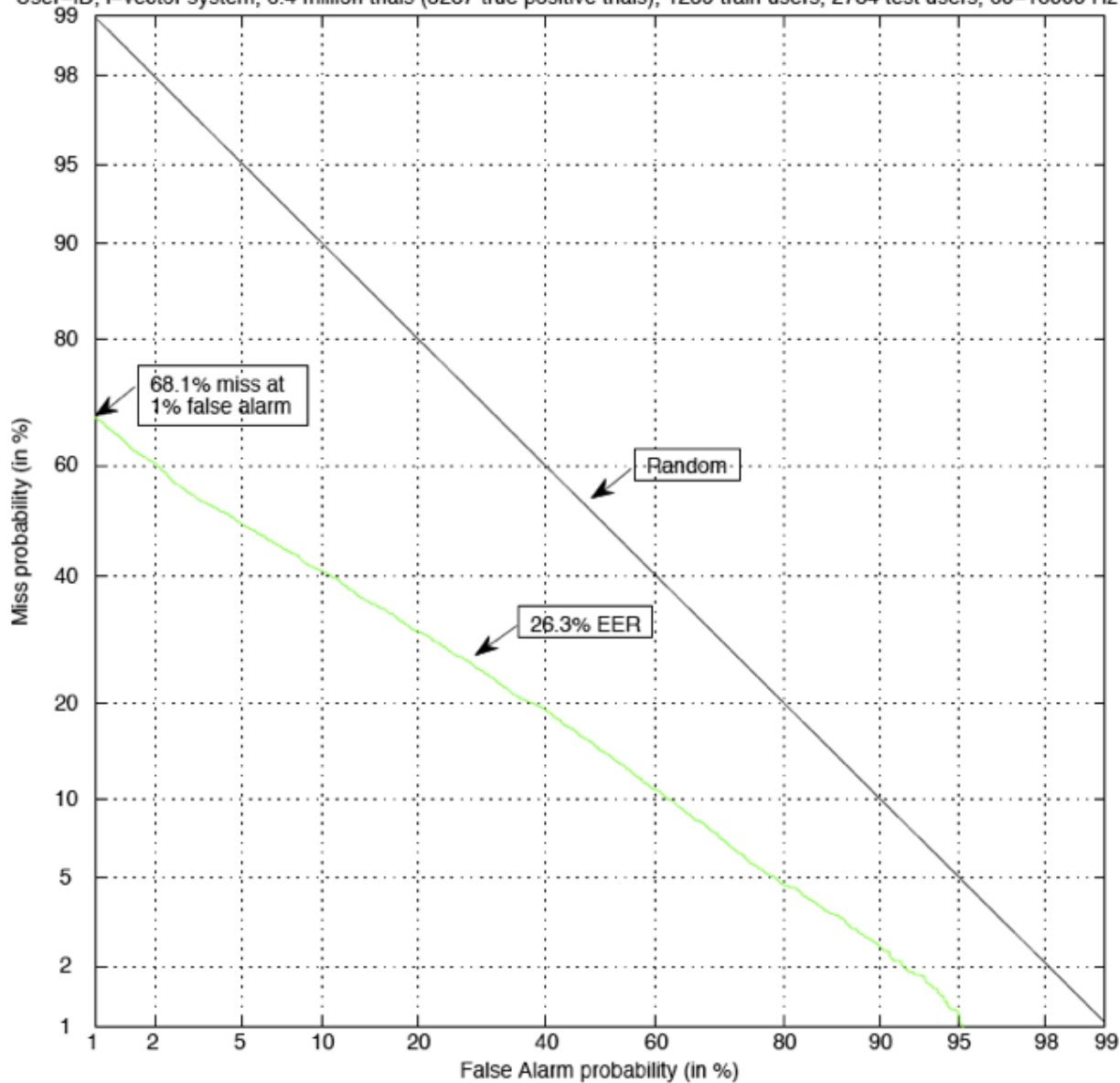
68.1% miss at 1% false alarm

Random

26.3% EER

Miss probability (in %)

False Alarm probability (in %)

Persona Linking using Internet Videos

Persona Linking using Internet Videos

Result:

H. Lei, J. Choi, A. Janin, and G. Friedland: "Persona Linking: Matching Uploaders of Videos Accross Accounts", at IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Prague, May 2011

Persona Linking using Internet Videos

Result:

On average having 20 videos in the test set leads to a 99% chance for a true positive match!

H. Lei, J. Choi, A. Janin, and G. Friedland: "Persona Linking: Matching Uploaders of Videos Accross Accounts", at IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Prague, May 2011

Question

Question

And now? What do you think has to be done?

Question

And now? What do you think has to be done?
a) Nothing can be done. Privacy is dead.

Question

And now? What do you think has to be done?

a) Nothing can be done. Privacy is dead.

b) We need to educate people about this and try to save privacy. (fight)

Question

And now? What do you think has to be done?

    a) Nothing can be done. Privacy is dead.

    b) We need to educate people about this and try to save privacy. (fight)

    c) I will really think before I post, and I agree with b).

Question

And now? What do you think has to be done?

    a) Nothing can be done. Privacy is dead.

    b) We need to educate people about this and try to save privacy. (fight)

    c) I will really think before I post, and I agree with b).

    d) I will really think before I post, and I agree with a).

Question

And now? What do you think has to be done?

a) Nothing can be done. Privacy is dead.

b) We need to educate people about this and try to save privacy. (fight)

c) I will really think before I post, and I agree with b).

d) I will really think before I post, and I agree with a).

e) I won't post anything anymore! (flee)

More examples and more discussion

More examples and more discussion

http://cybercasing.blogspot.com

What to do about it?

(under construction)

What to do about it?

http://teachingprivacy.icsi.berkeley.edu

(under construction)

Thank You!

Questions?

Work together with:
Robin Sommer, Jaeyoung Choi, Luke Gottlieb, Howard Lei, Adam Janin, Oriol Vinyals, Trevor Darrel, Dan Garcia, K. Ramchandran, E. Venketsan, and others.