Lecture 5: Technology – 3

Balachander Krishnamurthy

AT&T Labs–Research

`http://www.research.att.com/~bala/papers`

Mobile OSNs

- 3+ billion cellphones
- Many are “smart” (not remotely all)
- Primary means to access OSNs for many users (nearly half in FB)
- Eminently suited for Twitter
- Evolution over time: see Buongiorno’s white paper

[LWF] White paper Mobile social networking

Buongiorno white paper

Talks about 4 generations of evolution

1. '99–'00: Chat rooms, anonymous. Pre-pay, subscription-based
2. '04–'06: photos, mobile search, voting. Pre-pay, subscription-based
3. '08–'09: groups, alerts, games. Increasingly ad-supported
4. '10: can hide, video, multipoint chat, mobile gaming. With virtual currency

Examples of mobile OSNs

Many mOSNs did not exist prior to the widespread use of mobile devices

Brightkite, Buzzd, Dopplr, Foursquare, Gowalla, Gypsii, Loopt, Radar, Urbanspoon, Wattpad and Whrrl.

Gowalla was bought by FB and shut down. Unclear how many of the above other than Foursquare are active/popular

Bebo, Facebook, Hi5, Linkedin, Livejournal, MySpace, Twitter, Flickr, Yelp all have mobile interfaces: both for Web access and via apps (most common usage vector)

What's new in mOSNs?

- Notion of presence (are you on the mOSN now?)
- Location, location, location
- Unique device id (unless you are in .it)
- Shorter sessions, byte contribution drops due to device interface
- Sharper cultural differences due to varying degress of popularity of mobile devices
- Flat data rate plans leads to explosion of use
- And of course iPhones followed by Android
- Explosion in external apps (simpler in some sense)
- Increased privacy leakage vectors

Privacy Issues

Many mOSNs have a “check-in” mechanism—both establishes a user’s *presence* on the mOSN and the user’s current *location*.

Mobile devices typically have a unique device identifier, which is often used as verification for installing approved apps on a user’s mobile device.

If this unique identifier is leaked to a third-party via an application and can be associated with a user’s identity, this becomes a privacy problem.

New privacy concerns

- It is not uncommon to link mOSNs to traditional OSNs like FB, Twitter
- Information shared with a mOSN connected to a traditional OSN is also shared with that OSN
- Such a transitive closure of leakage is among the hardest to track
- Understudied and so a potential vector to examine in your project

Non-US centric OSNs

- Note that I did not cover all US-based OSNs (e.g., Google Plus)
- Path: Sara
- Chinese OSNs: Weibo, Renren, kaixin001.com, qq, wechat
- Iranian OSN?
- CyWorld, KakaoTalk (.kr)?
- Mixi (.jp)
- A brief presentation will be given by you!

Special purpose OSNs: spOSNs

- Pinterest
- SnapChat
- Instagram
- Last.fm
- WhatsApp
- Foodspotting
- Others?

Popular spOSNs: 1 – Pinterest [GBCT13]

- Tens of millions of users, rapid growth (grew 4000% in 2011!), vast fraction of users are women
- Goal: connect people through things they find interesting
- Can be images/photos/recipes etc.
- Pin an object which can then be repinned with usual semantics
- Oddly enough, geography appears to play little role in popularity of objects (although, as expected, being female helps)
- Written in Django Python (a web app framework) like Instagram or Mozilla
- Third largest OSN in the US (after FB, Twitter)! Valued at around 4B USD.
- Already used by scammers (Fake Starbucks coupons as phish)

[GBCT13]: I Need to Try This"?: A Statistical Overview of Pinterest

Popular spOSNs: 2 – Snapchat

- Started with a notion of protecting privacy!
- Content “vanishes” after some time
- Not indexed and thus not searchable even by “friends”
- Mimics transient conversations (or quick chats, hence the name)
- There has been speculation on whether this actually happens
- Fast growth among a desirable demographic (13-16 year olds)
- FB tried to buy it and failed
- A good choice for your project!

Popular spOSNs: 3 – Instagram

- Started 2010, acquired by FB in 2012
- Visibly recognizable photos–look like polaroid/instamatic shots
- Photo editing/sharing app highly popular with young users
- Instagram.com/username makes it easier to find people
- Some unusual account itssteviewonder, cashcats, rickpoon
- Accounts public or private (limited control)
- Can block individual users – only on IG, but can be seen elsewhere (e.g. if tweeted)

Popular spOSNs: 4

Last.fm

- Music oriented (quite old: started in 2002!)
- Recommend music of interest
- Connect similar minded aficionados
- Historical value: see your own music interests change over time

WhatsApp

- Private message sharing (you can pick contacts)
- Audio, video, photos
- Select sub-group to be in touch with
- Paid service (after a free trial period)

Privacy settings in OSNs

First study: Krishnamurthy and Wills [KW08]

- Notion of privacy bits and groups of bits in OSNs.
- User's privacy controls—what settings are available.
- Default settings and their implications.
- What users do with these settings.
- The role of third party domains in aggregating user-related data and contrast with traditional Web sites.
- Ideas for providing better privacy protection in OSNs.

[KW08] Characterizing privacy in online social networks, WOSN 2008

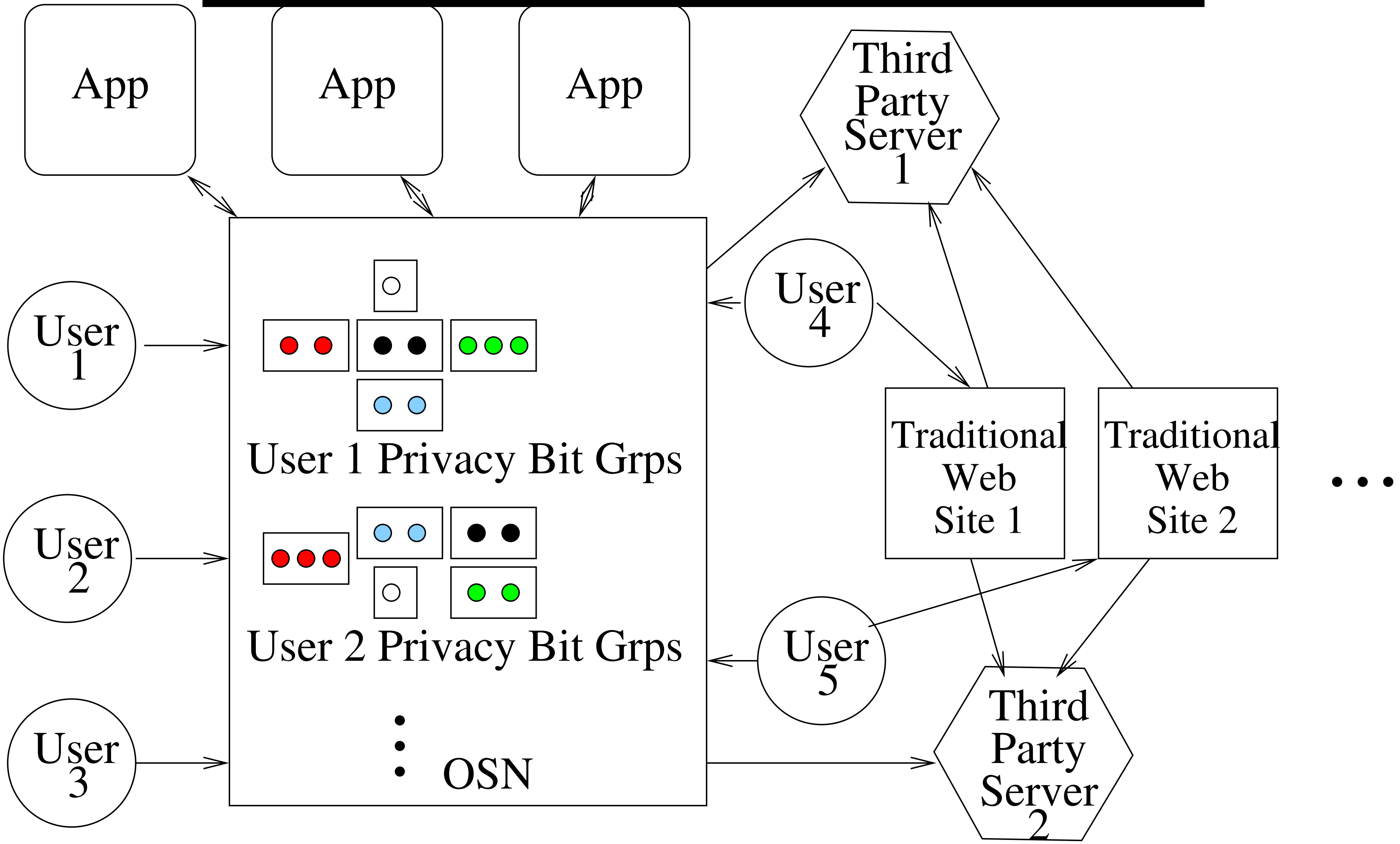
Privacy Bits

- A bit is a piece of private information
- There are many such bits; e.g., name, age, DOB, a friend, user uploaded content, comments etc.
- Some of the bits can be grouped. E.g.,

  Thumbnail - name, photo
  Greater profile - interests, relationship status etc.
  List of friends
- No way to vouch for the accuracy of these bits or even if the user is a real person.

.

Privacy Information and Potential Leakage



Privacy Groups and Entities

What can be shared:

- Can be ordered L to R by users based on their comfort level
- Groups on the right can be freely shared, on the left not so
- Left may be DOB, right may be name
- They can be stacked vertically if they are equally important

With whom:

OSNs let users to grant privileges to different entities

Typically: user, user's friends, all users

Some OSNs grant privileges to friends of friends or users in a common "network"

Facebook Privacy Settings

Privacy Bit Group	Self	Friends	Friends of Friends	Friends+ Networks	All
Thumbnail	-	○	○	○	$\boxed{\circ}$
Greater Profile	-	○	○	$\boxed{\circ}$	-
List of Friends	-	○	○	○	$\boxed{\circ}$
User Gen. Content	-	○	○	$\boxed{\circ}$	-
Comments	○	○	○	$\boxed{\circ}$	-

Key: Can be set: ○; Not possible: -; Default: $\boxed{\circ}$

Greater profile ('profile' in FB) can be set to be viewable by friends, FoF, or friends and users in same networks (default). By default thumbnail and list of friends is viewable by all.

Double vertical line in table is a threshold for user control

MySpace Privacy Settings

Privacy Bit Group	Self	Friends	Friends of Friends	Friends+ Age>18	All
Thumbnail	-	-	-	-	$\boxed{\circ}$
Greater Profile	-	○	-	○	$\boxed{\circ}$
List of Friends	-	○	-	○	$\boxed{\circ}$
User Gen. Content	-	○	-	○	$\boxed{\circ}$
Comments	-	○	-	○	$\boxed{\circ}$

Key: Can be set: ○; Not possible: -; Default: $\boxed{\circ}$

- Coarse-grained settings—all or nothing settings.
- Everyone has access to everything by default.
- Privacy controls for other popular OSNs (Bebo, Digg, Friendster, Hi5, Imeem, LiveJournal, Orkut, Twitter and Xanga) tend to be similar to coarse granularity and all-or-nothing settings of MySpace.

Use of Facebook Privacy Settings

Need to study networks: regional, high school, work

Anyone can join a regional network, others require some nominal ‘proof’ (an organization issued email address).

User can be in one regional network at a time.

506 regional networks (April 2008)

272 in US: cities and their region

234 global networks: cities in Canada/U.K., but countries elsewhere

We picked 20 U.S. and 18 non-U.S by first subdividing each set of networks into four size ranges then choosing specific networks within each range to ensure size and geographic diversity.

Used random network browsing feature of Facebook to obtain users within a network.

Privacy Settings in U.S. Facebook Regional Networks

Regional Network	Users (K)	%View Profile	%View Friends
New York,NY	866	53	78
Chicago,IL	649	54	78
Los Angeles,CA	595	62	82
Atlanta,GA	390	56	82
Dallas/FW,TX	336	63	84
Seattle,WA	210	64	83
Sacramento,CA	99	76	90
Des Moines,IA	83	67	85
Okla City,OK	80	71	87
Greenville,SC	66	72	90
Syracuse,NY	54	75	90
Worcester,MA	45	77	94
Peoria,IL	44	77	93
Boise,ID	36	83	96
Tupelo,MS	29	76	98
La Crosse,WI	25	71	94
Monroe,LA	21	79	98
Ithaca,NY	17	78	95
Abilene,TX	10	82	97
Casper,WY	6	84	99

Strong negative correlation between network size and percentage of users allowing profile and friends to be viewed.

Facebook Settings Inference

- Strong negative correlation between network size and user profile visibility. Likewise with viewing friends.
- Same with non-US users—true across cultures.
- Users (apparently) care more about profile info than list of friends.
- Facebook allows further user control of access to some information in a user's profile (e.g. viewing Wall comments). Consequently the privacy of Wall comments is further protected than the View Profile setting.

Use of Third-Party Domains

Performed session of typical actions for each OSN while recording the set of servers contacted for the content of each page.

Executive summary of results:

- Same entities involved as for traditional Web sites (comparison with prior work).
- Users think they are giving information about themselves to their OSN, but others are getting access to what users are doing.

Top Third-Party Domains in OSN Sessions

	Online Social Network					
Third-Party Domain	Fr'ster	Imeem	Bebo	Hi5	MySpace	Xanga
doubleclick.net	✓	✓	✓	✓	✓	✓
2mdn.net	✓	✓	✓	✓	✓	✓
advertising.com	✓	✓	✓	✓	✓	
atdmt.com	✓	✓			✓	
googlesyndication.com	✓		✓	✓	✓	
quantserve.com		✓		✓		✓
adbrite.com	✓	✓	✓			
google-analytics.com	✓	✓				✓
yieldmanager.com			✓	✓		✓

	Online Social Network				
Third-Party Domain	Digg	LiveJ	Facebook	Twitter	Orkut
doubleclick.net	✓	✓			
2mdn.net	✓				
advertising.com			✓		
atdmt.com	✓	✓	✓		
googlesyndication.com		✓			
quantserve.com	✓				
adbrite.com					
google-analytics.com				✓	
yieldmanager.com					

What Do We Propose?

- Break privacy into groups
- Let users order them in terms of importance for them and specify how far in the L-R spectrum they are willing to allow access by default
- Let applications and OSNs ask for bare minimum and never more than supremum
- If needed information is within default, access is transparent
- If more is needed, user is asked

Personally Identifiable Information leakage via OSNs

“Controversial” paper [KW09], led to 15 nanoseconds of fame..

- PII: information which can be used to distinguish or trace an individuals identity either alone or when combined with other public information that is linkable to a specific individual.
- Users provide various pieces of PII to OSNs
- These are often visible to more than just “friends”
- Third-party servers for aggregating user viewing behavior are prevalent in popular Web sites *and* OSNs.
- Key question answered: is PII belonging to any user being leaked to these third-party servers via OSNs?

[KW09] On the Leakage of Personally Identifiable Information Via Online Social Networks

Overview of study results

- PII leakages *do* occur.
- via HTTP headers sent to third-party aggregators.
- Most users on OSNs are vulnerable to having their OSN identity information linked with tracking cookies.
- Shared this information to all the OSNs studied so that they may make informed decisions regarding preventative measures/subscriber notification.
- Goal not a legal examination of privacy policies, but to bring a technical examination of the observed leakage to the community's attention and to propose means to prevent such leakage.

Consequences

1. With tracking cookies having been set and gathered for several years to track user visits to non-OSN sites as well, it is possible for third-party aggregators to associate identity with those past accesses.

2. As users on OSNs will continue to visit OSN and non-OSN sites, such behavior in the future is also liable to be linked with their OSN identity.

Aggregators claim they create profiles of users based on their Internet behavior, but do not gather or record PII. Although we do not know that aggregators are recording PII, paper shows that it is undeniable that information is available to them-either directly or indirectly via OSN identifiers.

Degree of availability of PII (to OSN users) in 12 OSNs

Piece of PII	Always Available	Available by default	Unavailable by default	Always Unavailable
Personal Photo	9	2	1	0
Location	5	7	0	0
Gender	4	6	0	2
Name	5	6	1	0
Friends	1	10	1	0
Activities	2	8	0	2
Photo Set	0	9	0	3
Age/Birth Year	2	5	4	1
Schools	0	8	1	3
Employer	0	6	1	5
Birthday	0	4	7	1
Zip Code	0	0	10	2
Email Address	0	0	12	0
Phone Number	0	0	6	6
Street Address	0	0	4	8

Entries are counts of OSNs; columns go from bad to good wrt privacy concerns.

Leakage Detection Methodology

Used Live HTTP Headers extension for Firefox browser to capture complete HTTP header information while interacting with each of 12 OSNs studied.

In each case examined if and how OSN identifier is leaked to third-party aggregators.

Sample leakage (via an embedded object on myspace.com page):

```
GET /pagead/test domain.js HTTP/1.1
Host: googleads.g.doubleclick.net
Referer: http://profile.myspace.com/index.cfm?
fuseaction=user.viewprofile&friendid=123456789
Cookie:id=2015bdfb9ec||t=1234359834|et=730|cs=7aepmsks
```

Source of leakage

- OSNs assign unique IDs for their users that may be displayed as part of URL when user navigates around the OSN
- If the ID stays *within* the OSN, it is not a problem
- However, ID is ‘leaked’ to multiple outsiders, including 3d-party aggregators
- The ID, in conjunction with the aggregator’s tracking cookie leads to the actual privacy leakage
- The *same* tracking cookie is sent to the aggregator when the user visits other sites that trigger connections to the aggregator

Linkage is rearing its head now.

Technical manners of leakage

Three broad categories of leakage

1. OSN identifier (pointer to PII) via HTTP headers

2. OSN identifier through external applications

3. Raw bits of PII

Additionally, linkages across OSNs and non-OSNs are possible.

What can aggregators do with PII

- Tracking cookie from any other site is trivially *linkable* with OSN user

- Visits to non-OSN websites in the *past* and *future* can be linked with the information

- Searches are identifiable potentially with a person (assuming OSN ID is not falsified)

ID is leaked *in context* allowing lazy aggregation of data.

Protection Against PII Leakage

Parties:

1. User: Could filter out HTTP headersfiltering of cookies is already supported by browser.

Potential problem with the Referer header to leak private information has been known since 1996.

2. Aggregatorsfilter out PII-related headers. Make cookie semantics more visible.

3. OSNscould have strong default privacy protection. Easiest is to strip internal user identifier or map user identifier to opaque string on a per-session basis.

4. External applicationscould employ one of methods to strip the id or internally remap it.

References

[EGC$^+$10] W Enck, P Gilbert, B Chun, L Cox, J Jung, P McDaniel, and A Sheth. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. https://www.usenix.org/legacy/events/osdi10/tech/full_papers/Enck.pdf, 2010.

[GBCT13] Eric Gilbert, Saeideh Bakhshi, Shuo Chang, and Loren Terveen. "i need to try this"?: A statistical overview of pinterest. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2427–2436, 2013.

[Kri09] B. Krishnamurthy. A measure of online social networks. 2009. Invited paper.

[KW08] Balachander Krishnamurthy and Craig E. Wills. Characterizing privacy in online social networks. In *Proceedings of the first workshop on Online social networks*, 2008.

[KW09] B. Krishnamurthy and C. Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the Workshop on Online Social Networks*, August 2009. http://www.research.att.com/~bala/papers/wosn09.pdf.

[KW10] Balachander Krishnamurthy and Craig E. Wills. Privacy leakage in mobile online social networks. In *Proceedings of the Workshop on Online Social Networks*, June 2010. http://www.research.att.com/~bala/papers/pmob.pdf.

[LWF] Nick Lane and Nicky Walton-Flynn. White paper – mobile social networking. http://www.telecoms.com/files/2009/05/buongiorno_final-fmt_nl-3110-f.pdf.

[Mar10] Marco Balduzzi, et al. Abusing social networks for automated user profiling. In *RAID*, 2010. http://www.iseclab.org/papers/raid2010.pdf.

[WCG+11] David Wetherall, David Choffnes, B. Greenstein, Seungyeop Han, Peter Hornyack, Jaeyeon Jung, Stuart Schechter, and Xiao Wang. Privacy revelations for web and mobile apps. In *Proceedings of HotOS*, May 2011. `http://appanalysis.org/jjung/jaeyeon-pub/hotos2011-revelations.pdf`.