



Mobile App Privacy:

How Bad Is It & What Can We Do About It?

Norman Sadeh

Professor, School of Computer Science

Director, Mobile Commerce Lab.

Co-Director, MSIT in Privacy Engineering Program

Carnegie Mellon University

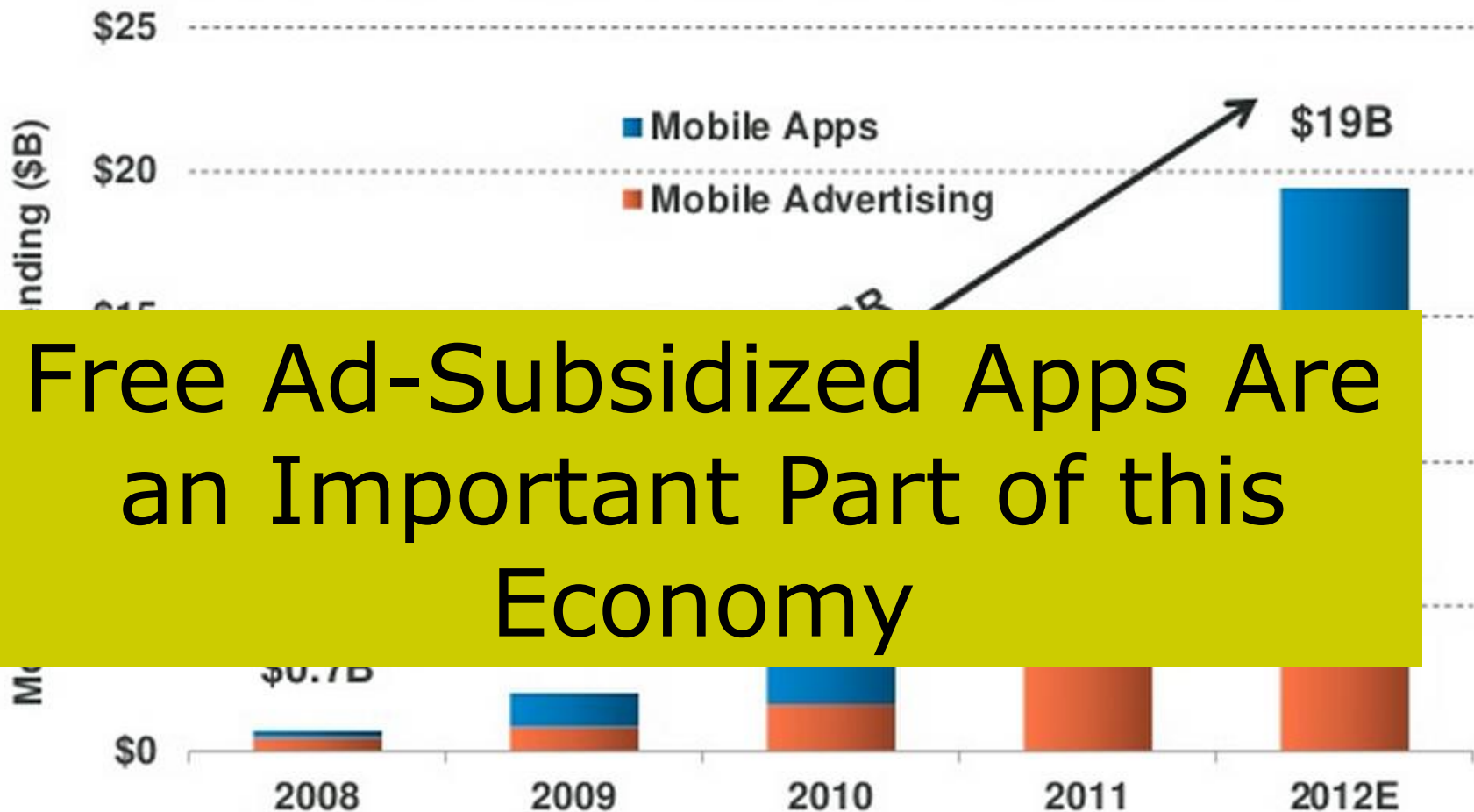
www.cs.cmu.edu/~sadeh --- sadeh@cs.cmu.edu



Outline

- The Mobile App Economy
- Mobile App Privacy
- What is being collected and for what purpose
- Why are current solutions ineffective?
- Tension between privacy and usability
- Some Promising Research Results
- Concluding Remarks

The Mobile App Economy (US)



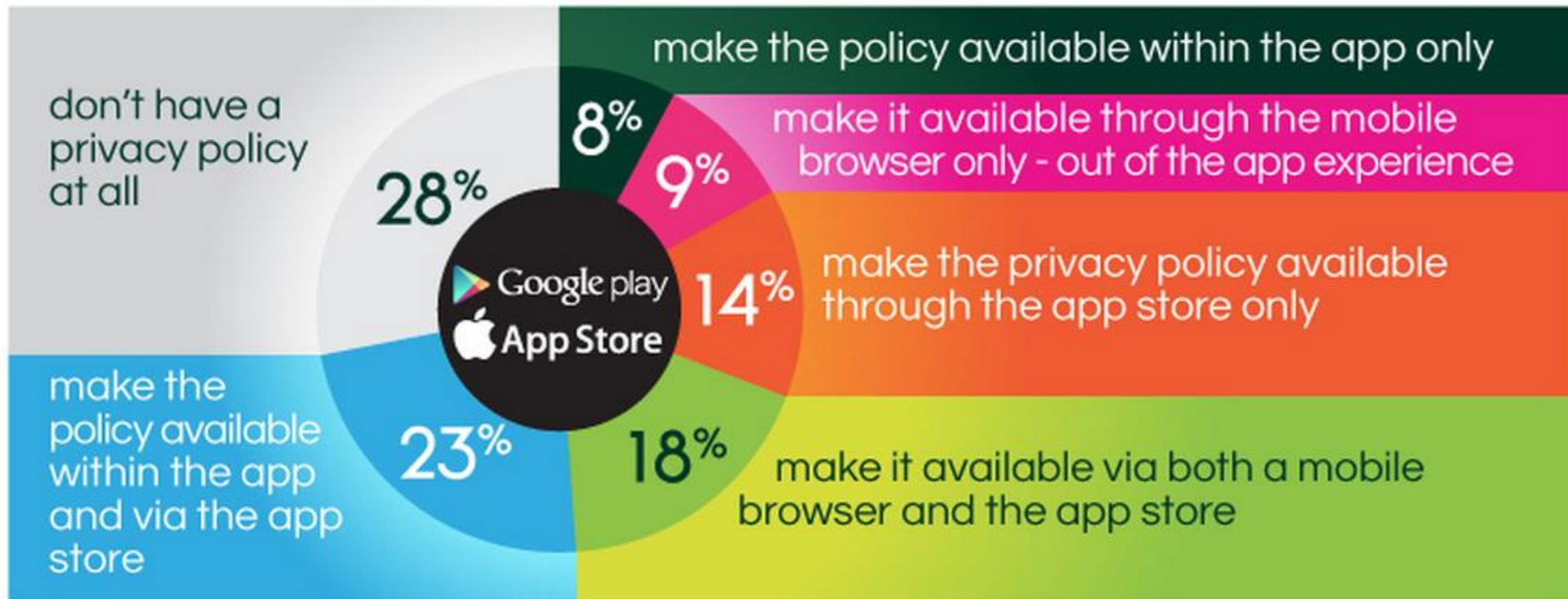
KPCB

Source: Gartner, eMarketer, Strategy Analytics. CAGR is compound annual growth rate.
Note: Apple has paid \$6.5B+ to developers as of 9/12, implying gross app market revenue of \$9B+ in 4 years; Google indicated during Q3:12 earnings call that its mobile revenue (from advertising and apps / content) run rate is \$8B+, up from \$2.5B mobile ad revenue run rate in Q3:11.

API Competition

- ❑ App Stores & Operating Systems compete for developers by exposing a growing collection of APIs...
 - ❑ Support new functionality/usage scenarios
 - ❑ Generate revenue

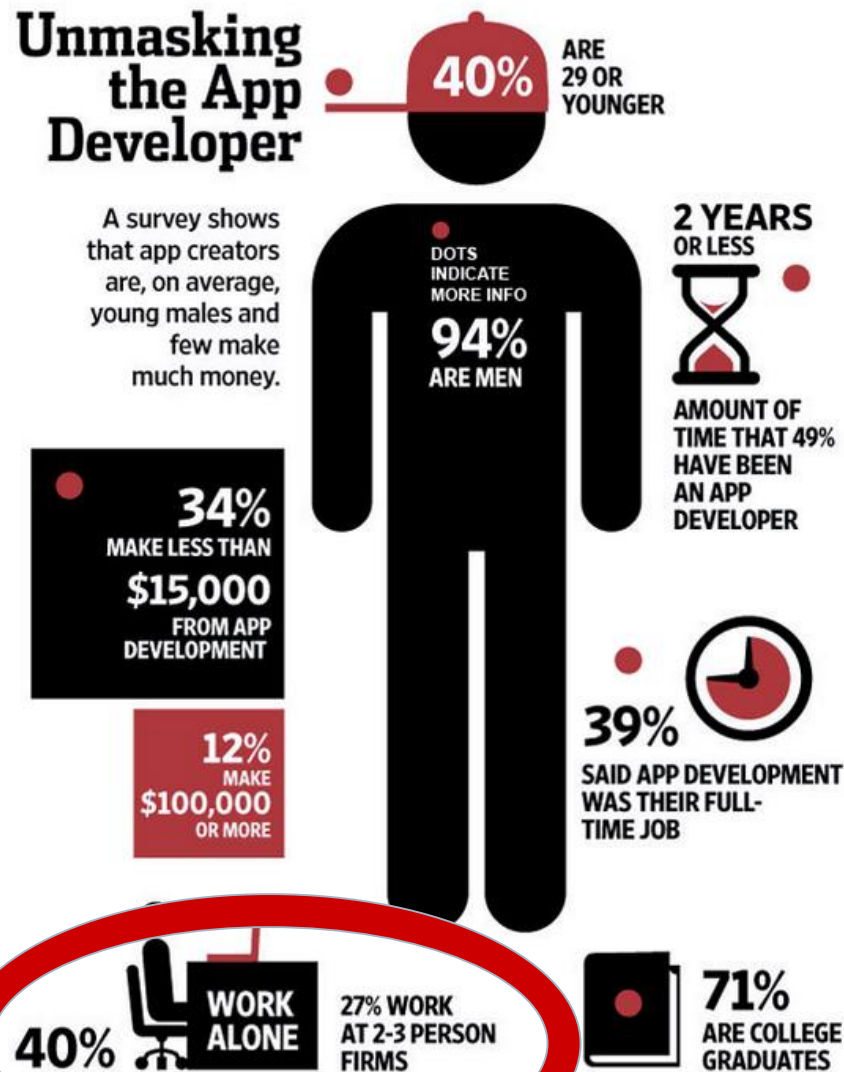
Where's the Privacy Policy?



Source: MSF – August 2013 --- <http://mefminute.com/2013/08/20/infographic-more-than-a-quarter-of-the-top-100-free-mobile-apps-dont-have-a-privacy-policy/#>

"2 (or 3) Guys in their Garage"

Unmasking the App Developer



Source: GigaOm Pro web-based survey with 352 respondents, Jan. 2012
by Alberto Cervantes/The Wall Street Journal

A Few Questions

How many of you would feel comfortable sharing the following information with your car insurance company?

- How many **miles/year** you drive

**Note that a smartphone
can help collect all this
information...today**

- ☐ including history of possible substance abuse
- How many hours you **sleep each night**
 - ☐ based on sensors
- ☐ What if insurance companies **purchased** some of this information from a **data broker** to decide whether to offer you coverage and what rate to quote?

Apps Often Collect More than They Need



Pandora gathers location, gender, year of birth, etc.



Path uploads entire contact list without user full consent.



Brightest Flashlight requires full Internet access, location, etc.

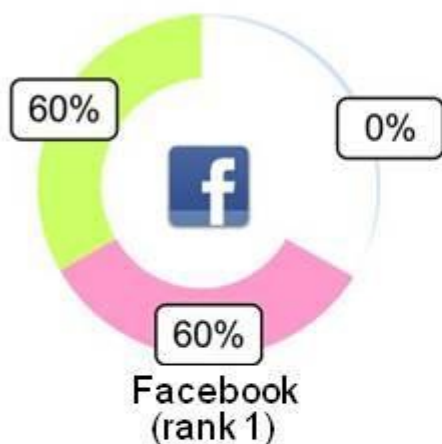


Bible accesses location.

People's Response When They Find Out...

Percentages of people surprised by an App's Permission Requests

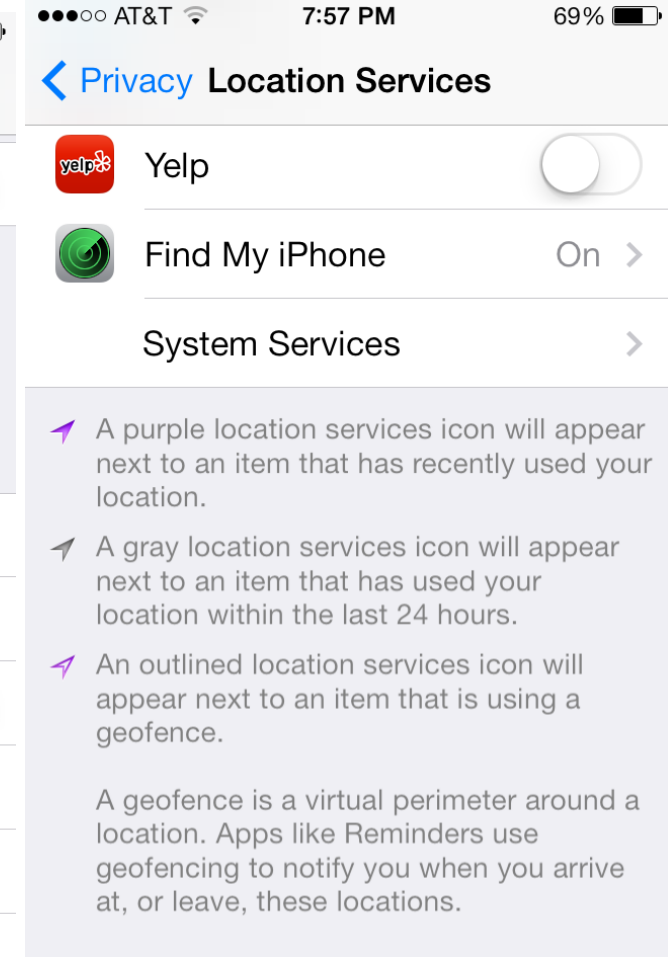
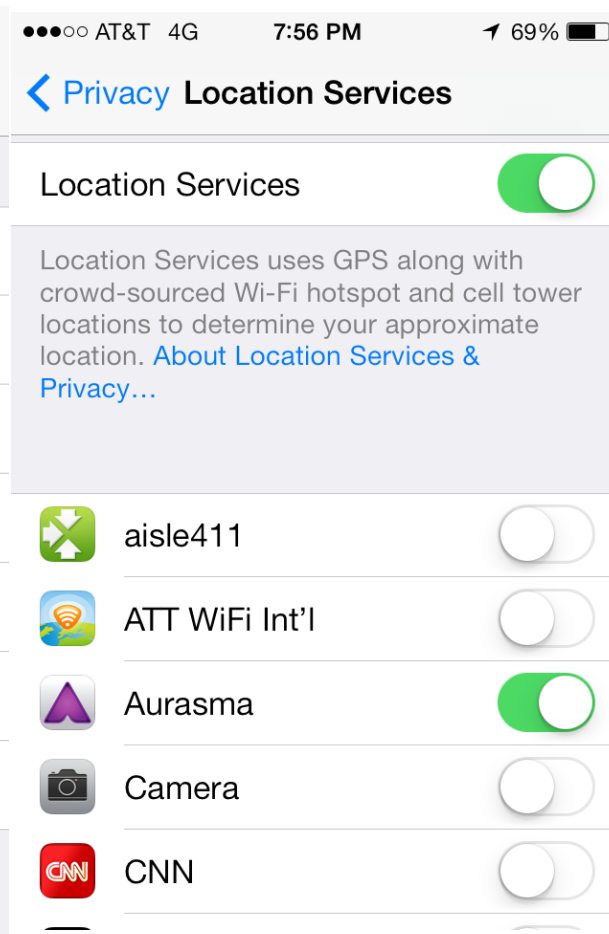
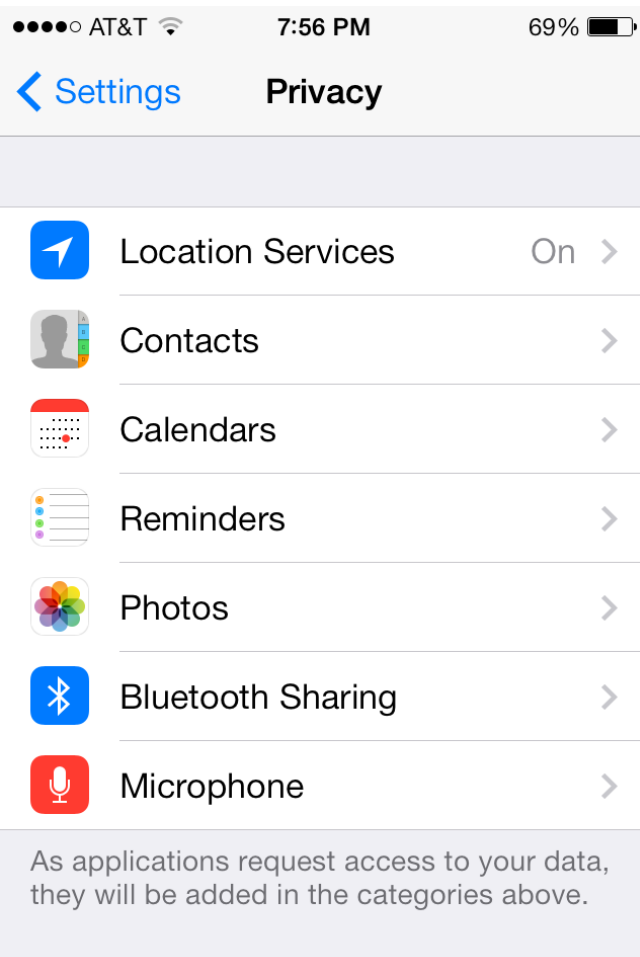
Location
Device ID
Contact List



J. Lin, S. Amini, J. Hong, N. Sadeh, J. Lindqvist, J. Zhang, "Expectation and Purpose: Understanding Users' Mental Models of Mobile App Privacy through Crowdsourcing", Proc. of the 14th ACM International Conference on Ubiquitous Computing, Pittsburgh, USA, Sept. 2012

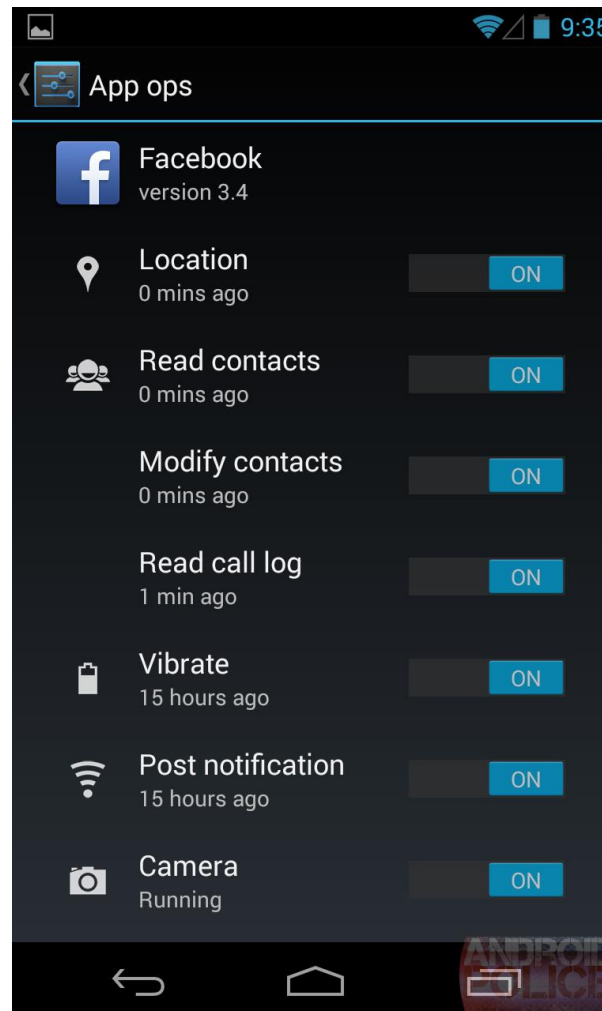
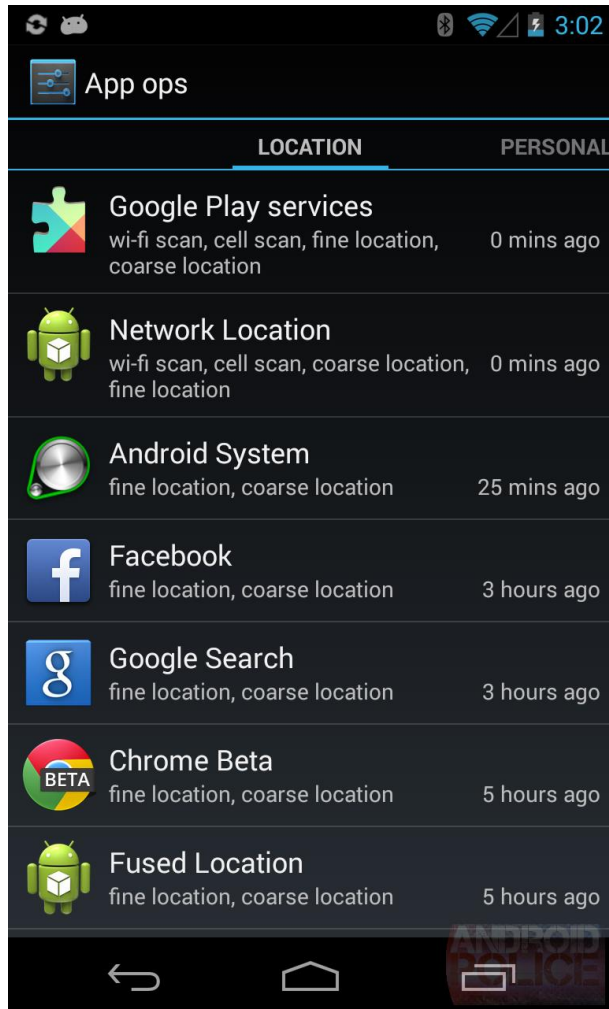
Explosion in Number of Privacy Settings (iOS)

A few iOS7 privacy screens – not even all location permission screens!



Explosion in Number of Privacy Settings (Android)

Android App Ops Hidden Permission Manager (introduced in Android 4.3 but dropped in Android 4.4...**not usable**)



Are All these Settings Good or Bad?

- ❑ In many ways, iOS and Android deserve credit for adding these settings
- ❑ Yet, the **number of settings users are expected to manage is unrealistically high**
 - A user with 40 apps and 3 permissions per app would have to manage 120 permissions!
- ❑ The information provided **does not explain how sensitive data/functionality is used**
 - Issue of **purpose** is central to people's preferences

Could a deeper understanding of people's privacy preferences help simplify decisions users have to make?

Two Parts to this Analysis

- ❑ Identifying the purpose of app permissions
- ❑ Collecting people's privacy preferences, taking into account purpose information

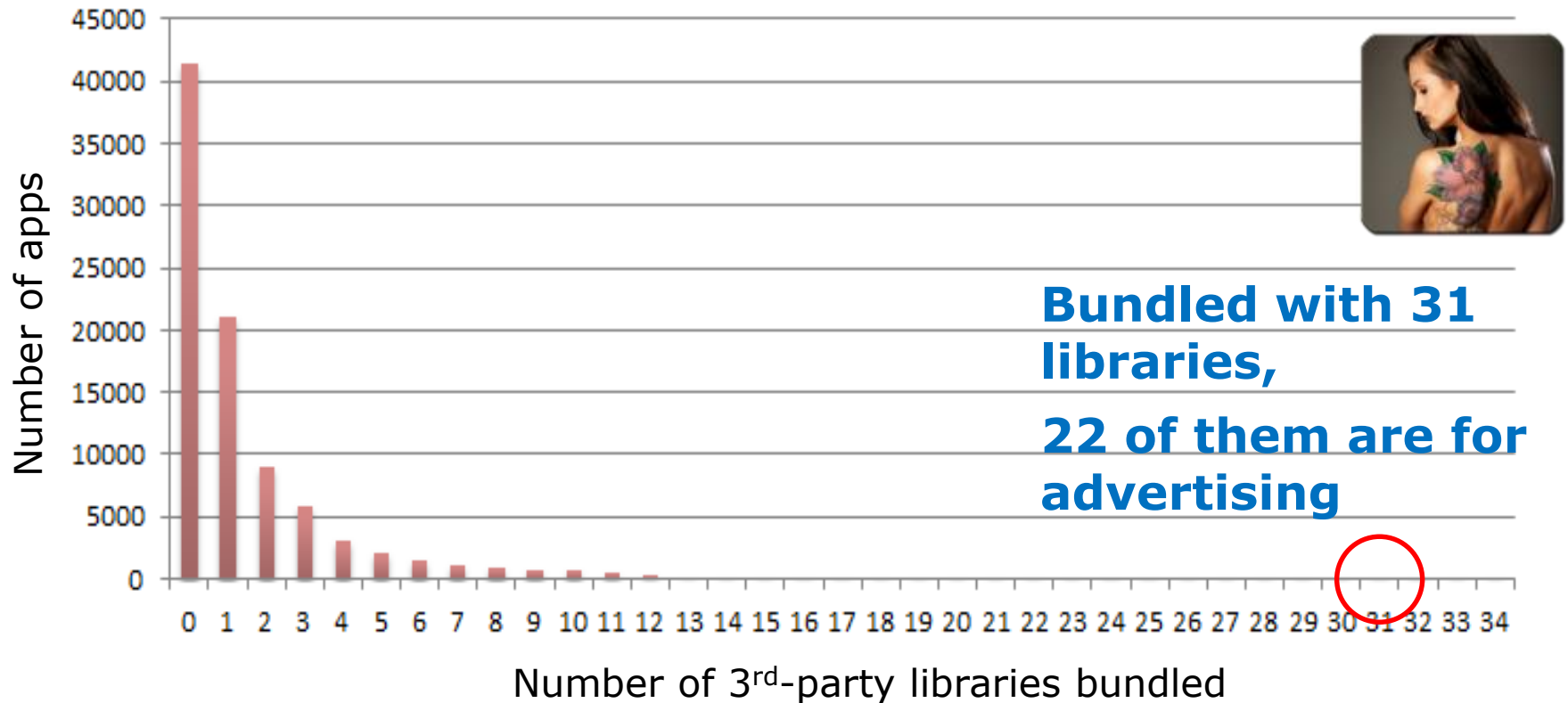
Identifying the Purpose of Permissions

- ❑ Androguard Reverse Engineering Tool
- ❑ Amazon EC2
 - 2035 instance hours → 1.23 minutes/app
- ❑ 89,903 apps successfully analyzed
 - 83.05% successful rate
 - Failures primarily due to code obfuscation



J. Lin, B. Liu, J.I. Hong, and N. Sadeh, "Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings", 2014 ACM Symposium on Usable Security and Privacy (SOUPS 2014), July 2014.

Uses of 3rd-party Libraries



Mean=1.59, SD=2.82, Median=1

Categories of 3rd-Party Libraries (Top 400)

Targeted Ads

SNS

**Customized UI
Component**

Game Engine

Payment

Mobile Analytics

Utility

Content Host

Secondary Market

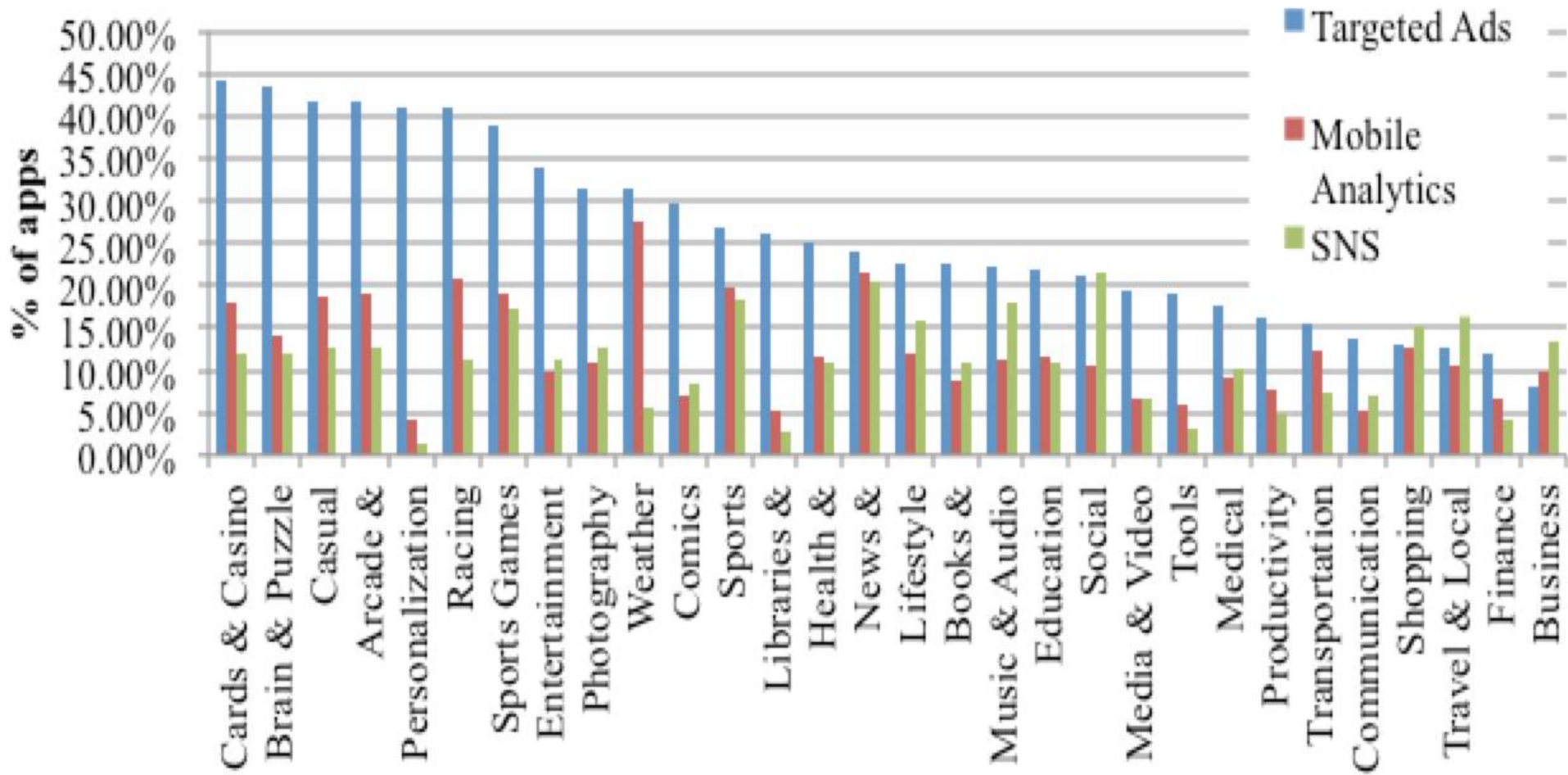
Permission Breakdown by Purpose

	Internal Use	Targeted Ads	Mobile Analytics	SNS
INTERNET	41.33%	47.48%	20.71%	16.30%
LOCATION	17.48%	72.94%	26.08%	6.07%
PHONE_STATE	24.55%	74.40%	16.04%	6.35%
READ_CONTACTS	52.07%	45.76%	-	2.81%
BLUETOOTH	86.54%	-	-	-
SMS	63.33%	38.81%	-	1.19%
GET_ACCOUNTS	32.51%	4.95%	-	8.04%
CAMERA	30.06%	17.45%	-	-
RECORD_AUDIO	91.91%	9.51%	-	-

In other words, **72.94% of apps requiring access to your location, use it for targeted ads.**

Some apps require some permissions for multiple purposes

Prevalence of 3 Types of 3rd Party Libraries

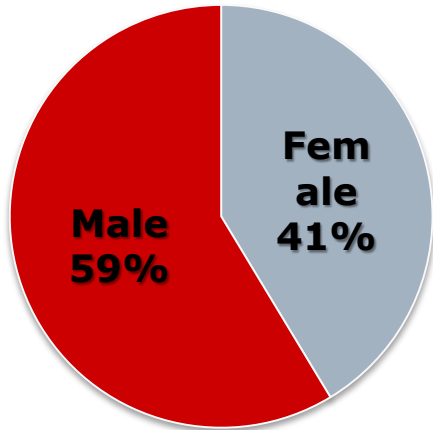


Understanding People's Privacy Preferences

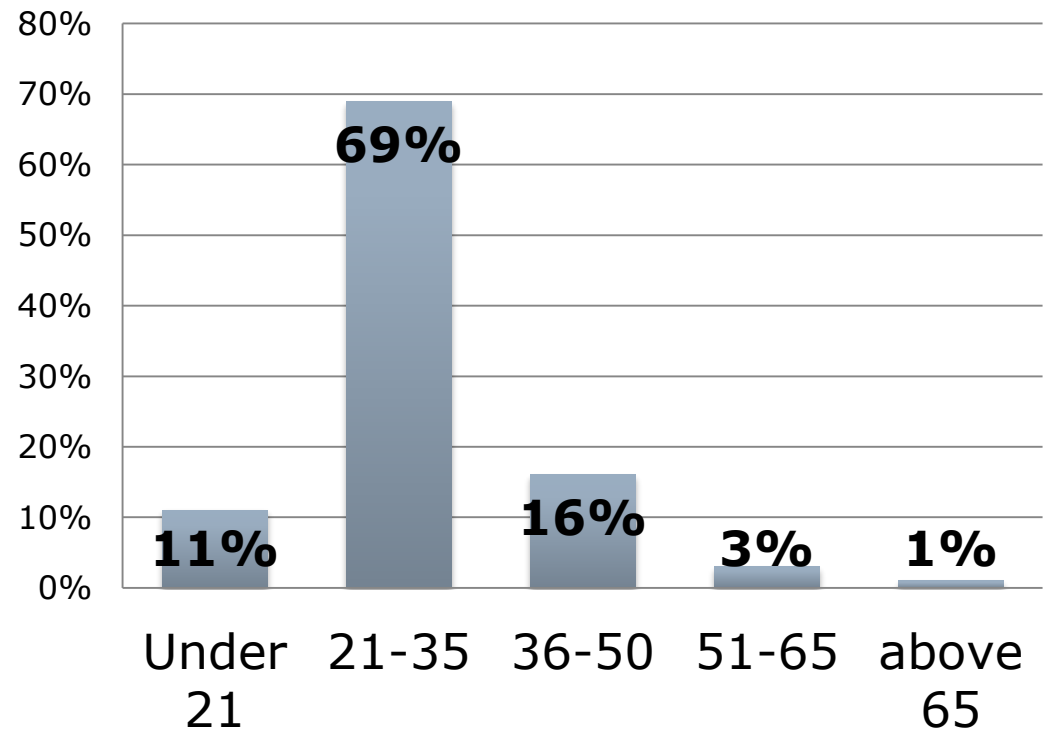
- ❑ Amazon Mechanical Turk, June 15 – June 30, 2013
- ❑ Total **725 U.S. smartphone users** participated
- ❑ 1200 HITs regarding to **837 mobile apps**
 - One HIT talks about one app, one permission, one purpose triple
- ❑ We eliminated HITs with fewer than **15 responses** --> total **21, 657 responses**

Participants

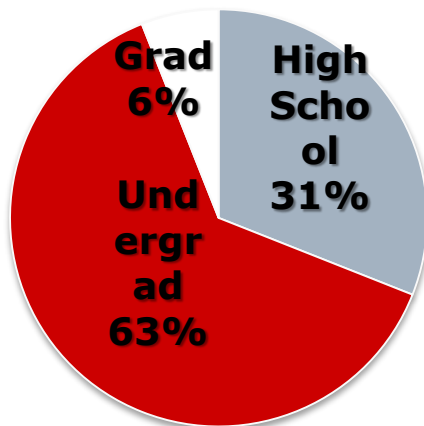
(a) **Gender**



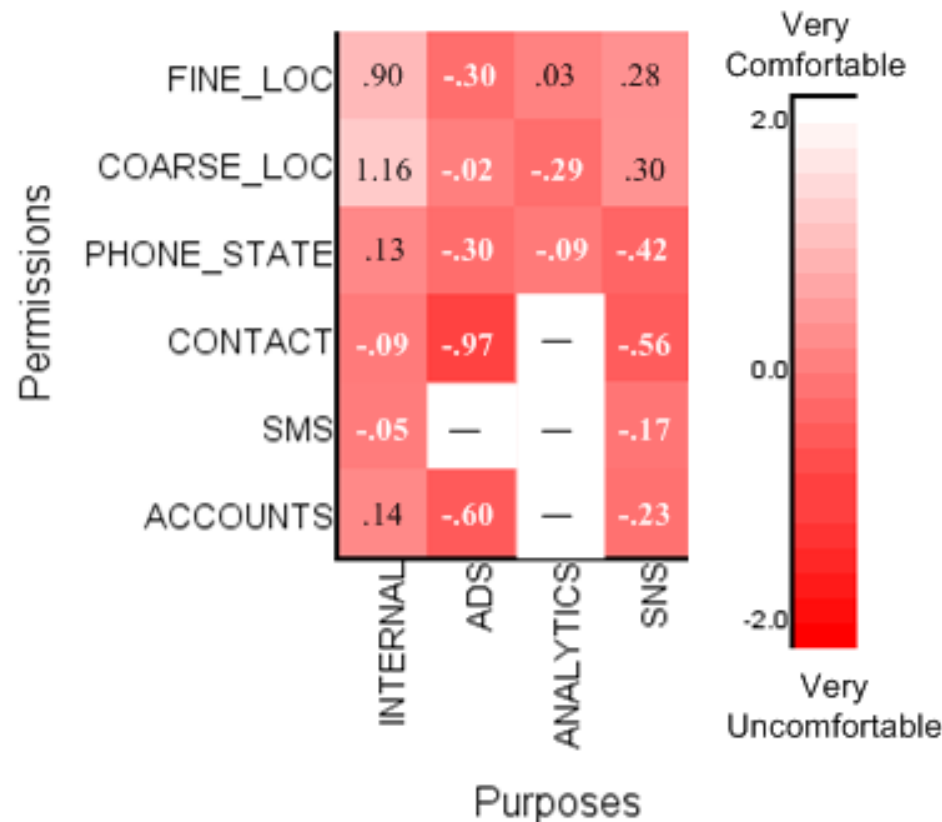
(b) **Age**



(c) **Education**



Android Permissions: Purpose Matters!



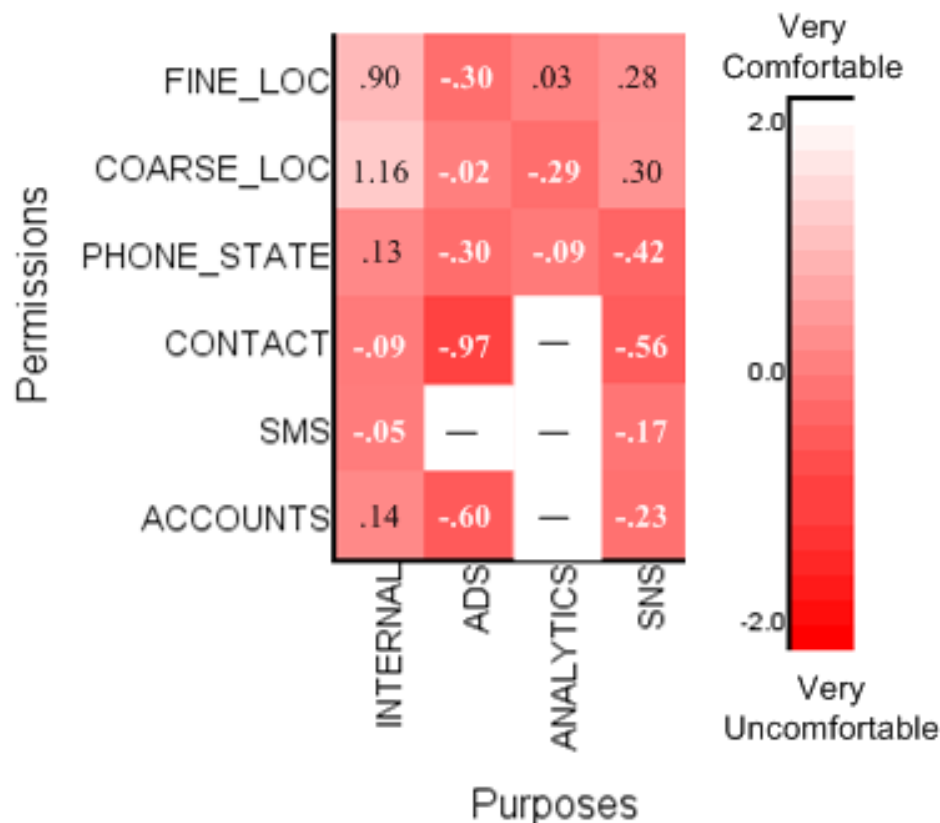
Users' Average Preferences

White → comfortable

Red → uncomfortable

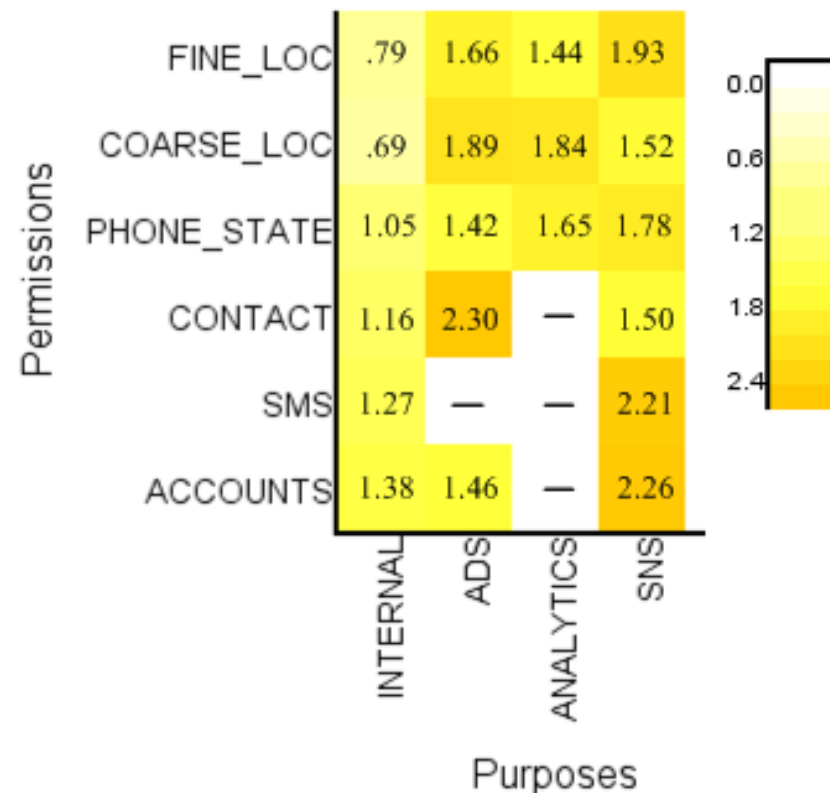
J. Lin, B. Liu, J.I. Hong, and N. Sadeh, "Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings", 2014 ACM Symposium on Usable Security and Privacy (SOUPS 2014), July 2014.

One Size-Fits-All Defaults Won't Work



Users' Average Preferences

White → comfortable
Red → uncomfortable



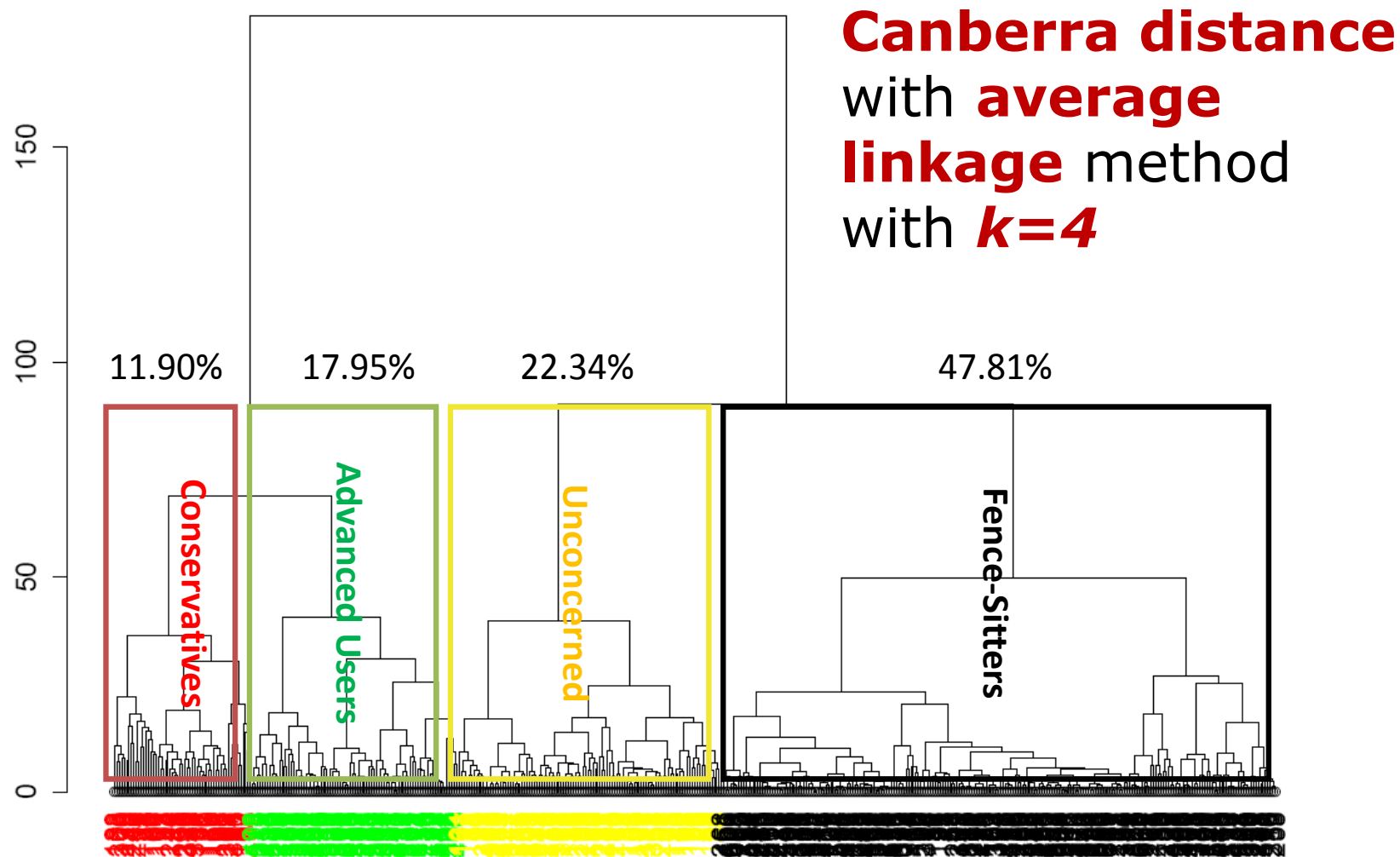
Variance among Users

Darker yellow → larger variance

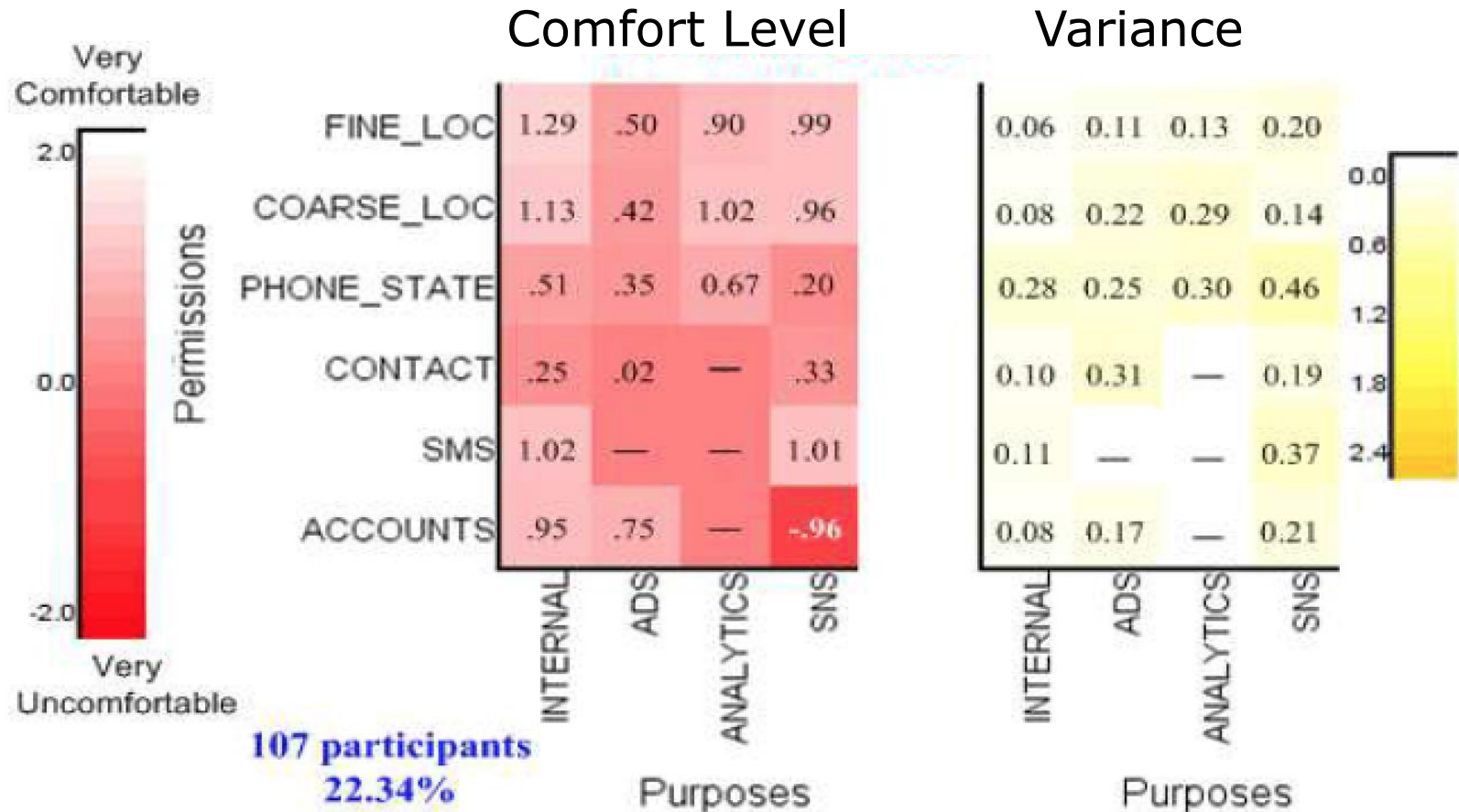
Tension between Privacy and Usability

- ❑ Unrealistically high number of settings
- ❑ Users cannot be expected to manage this level of complexity

Hierarchical Clustering

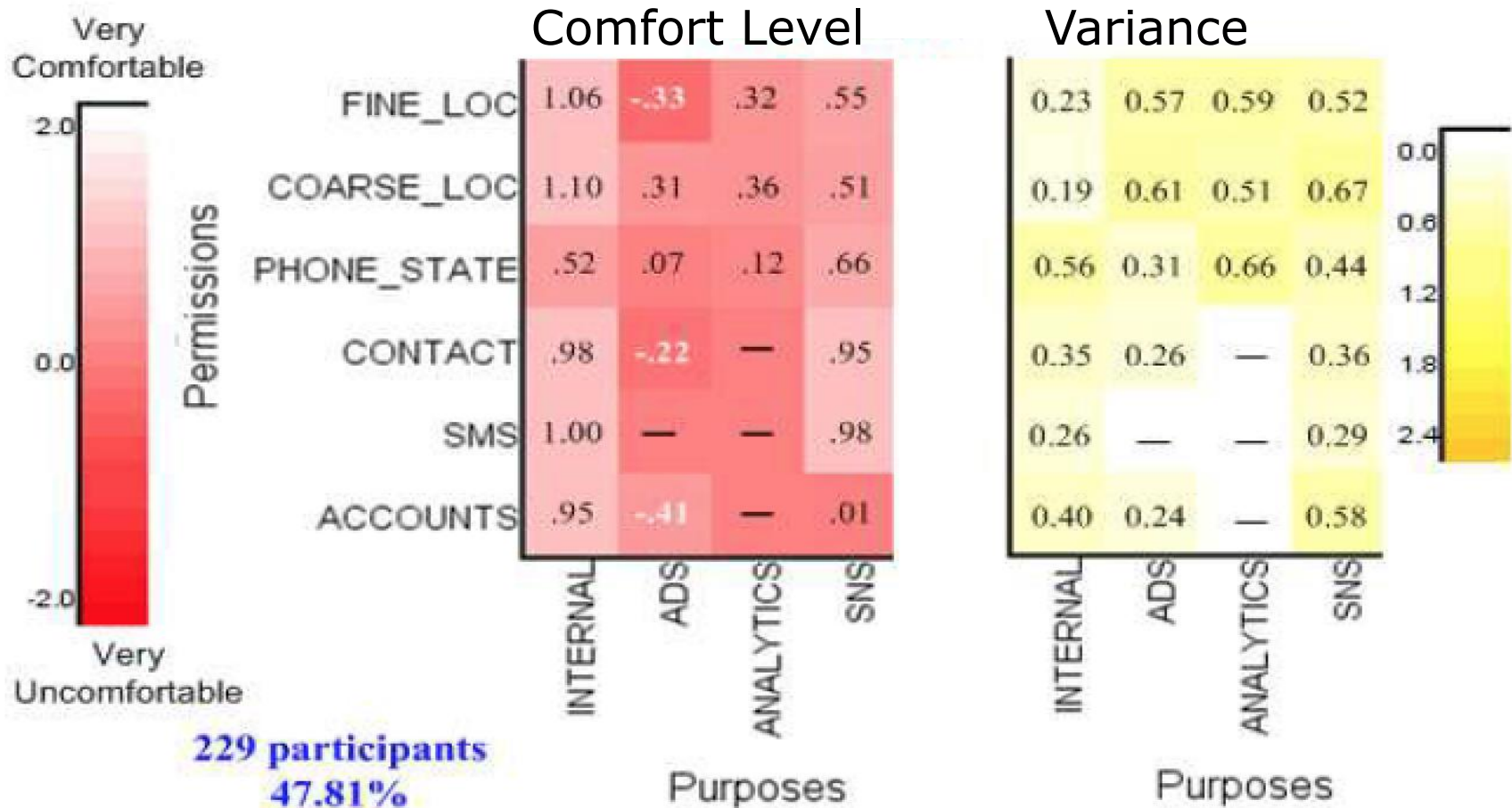


Unconcerned Cluster



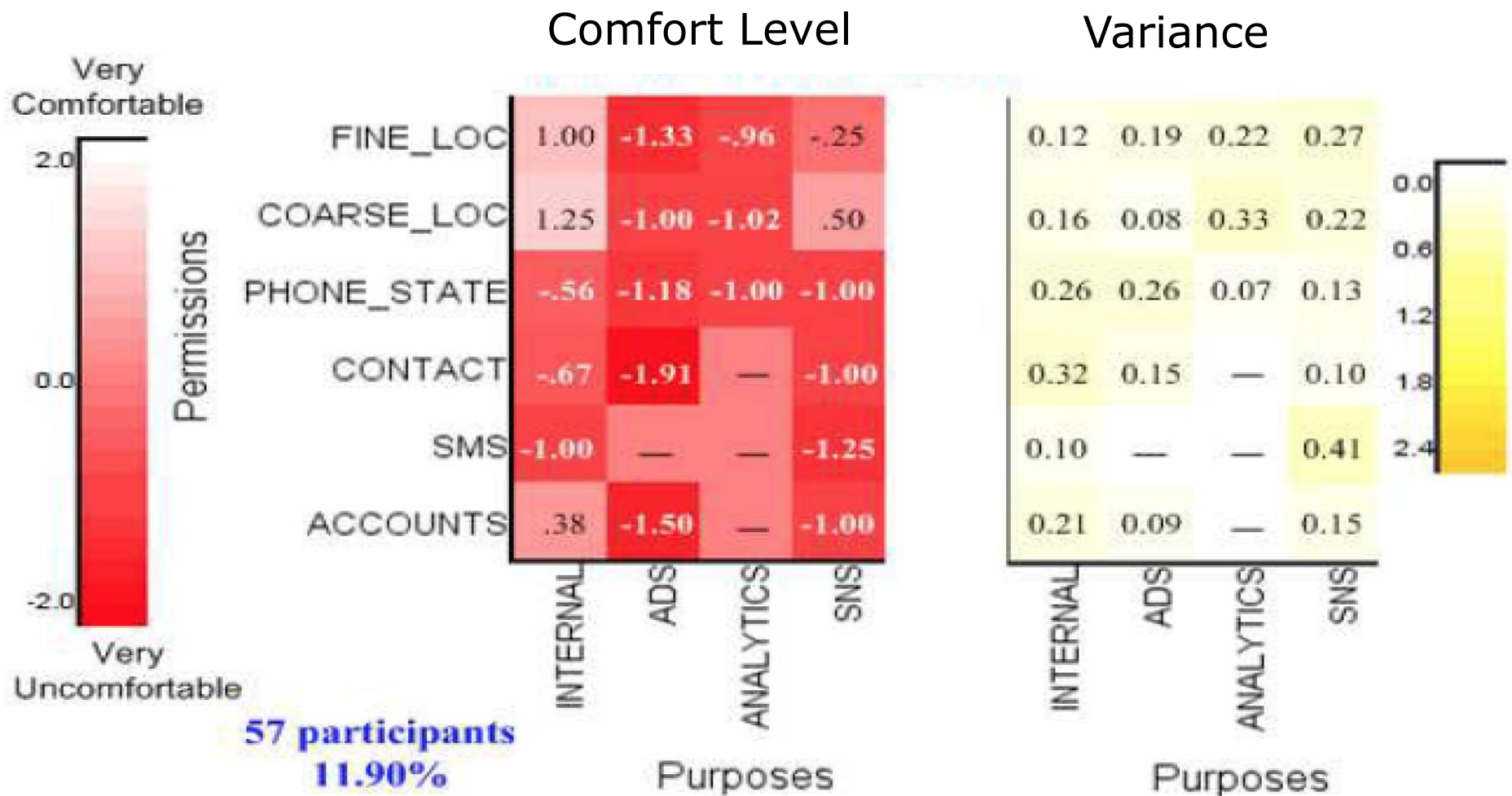
- Generally open to all types of disclosures
- Red in SNS/Accounts is probably a fluke – insufficient data

Fence-Sitters



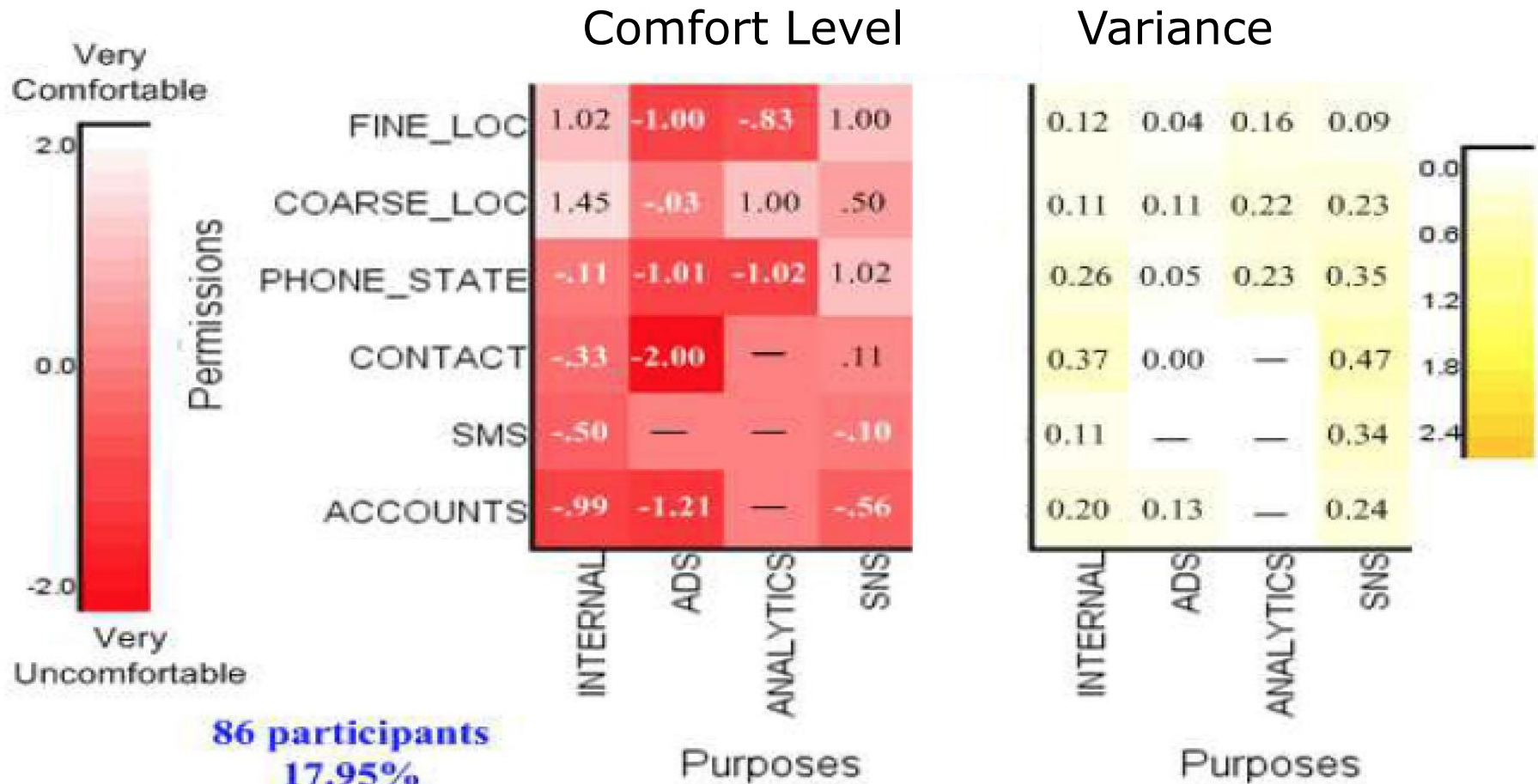
- Largest group of users (47.81%)
- Seem to have relatively neutral attitudes – could be habituation

Conservatives



- Uncomfortable letting external libraries access their information in general
- Even for internal purposes in the case of contact list, SMS and phone state

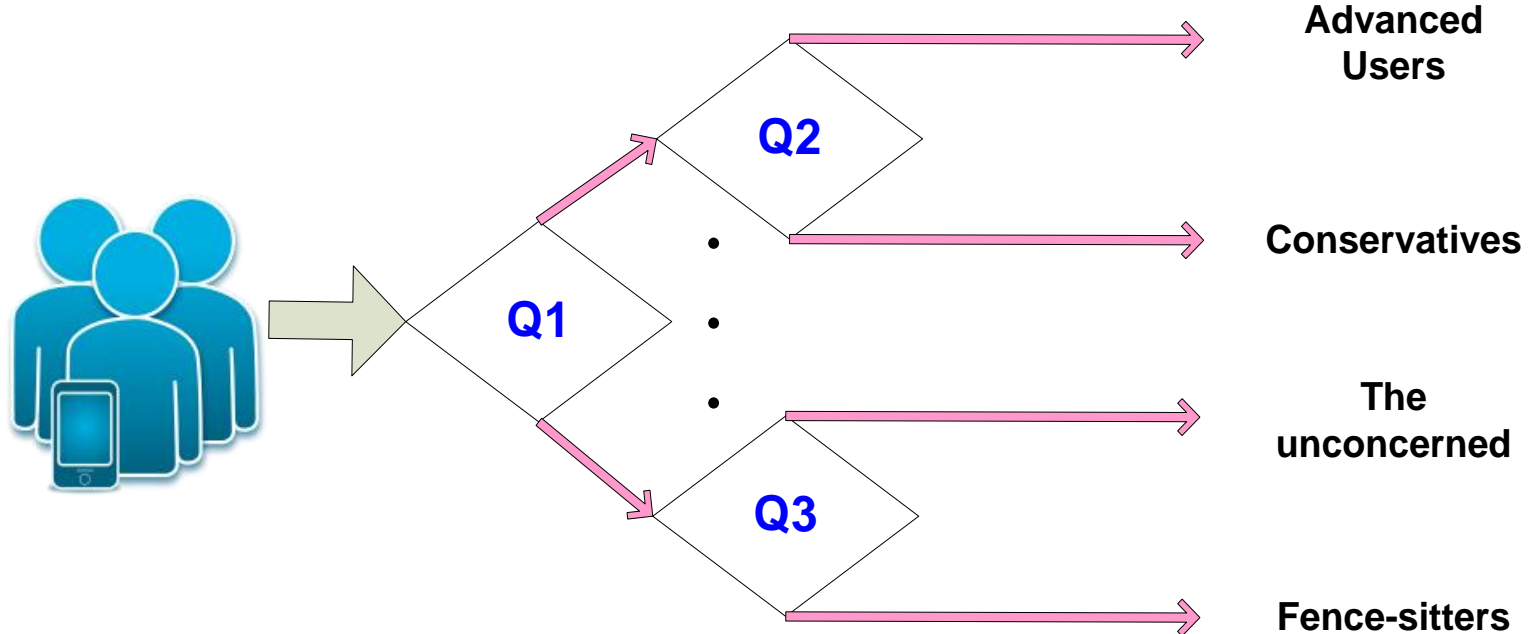
Advanced Users



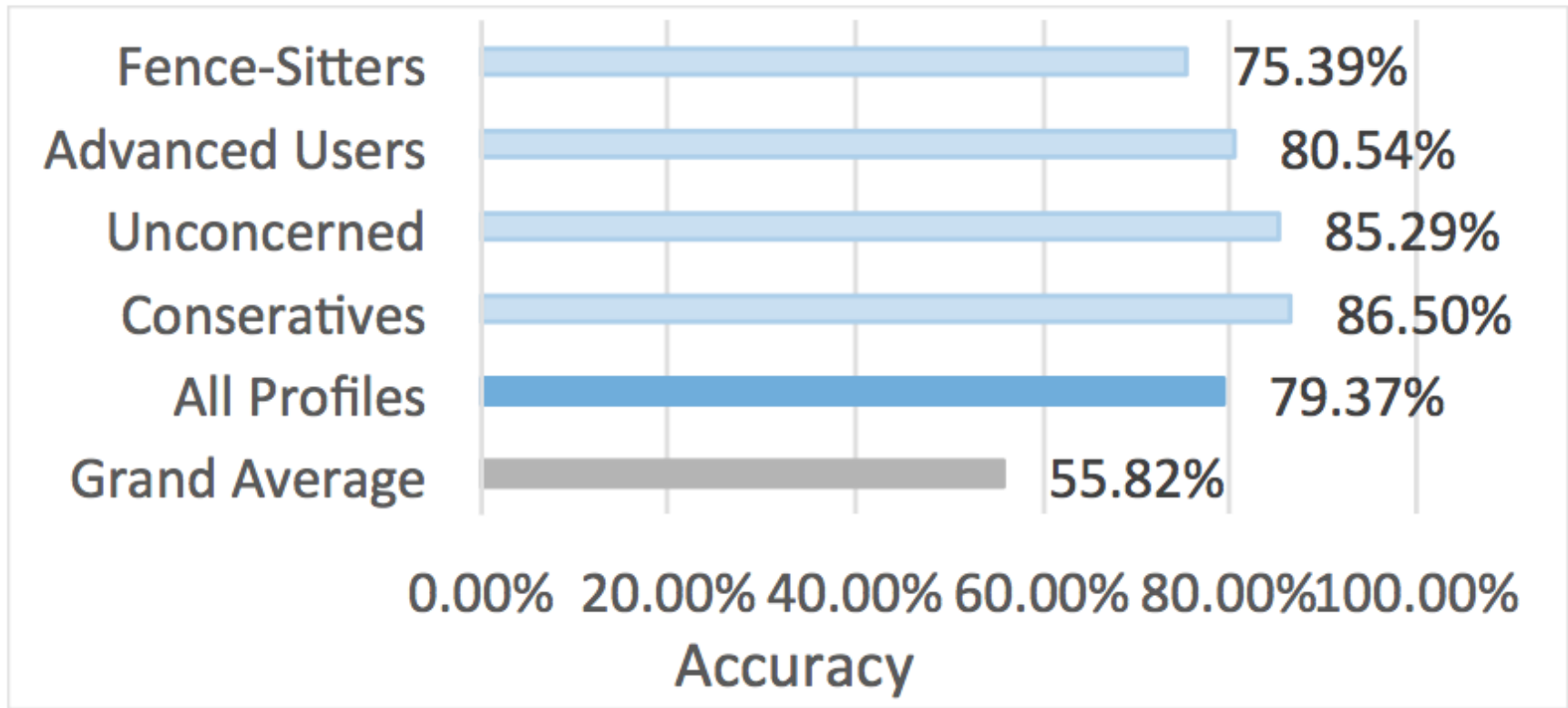
- Don't like ads and mobile analytics
- OK disclosing coarse location information, more cautious with fine location
- OK disclosing fine location with SNS

Identifying a User's Privacy Profile

- Asking users a small set of questions



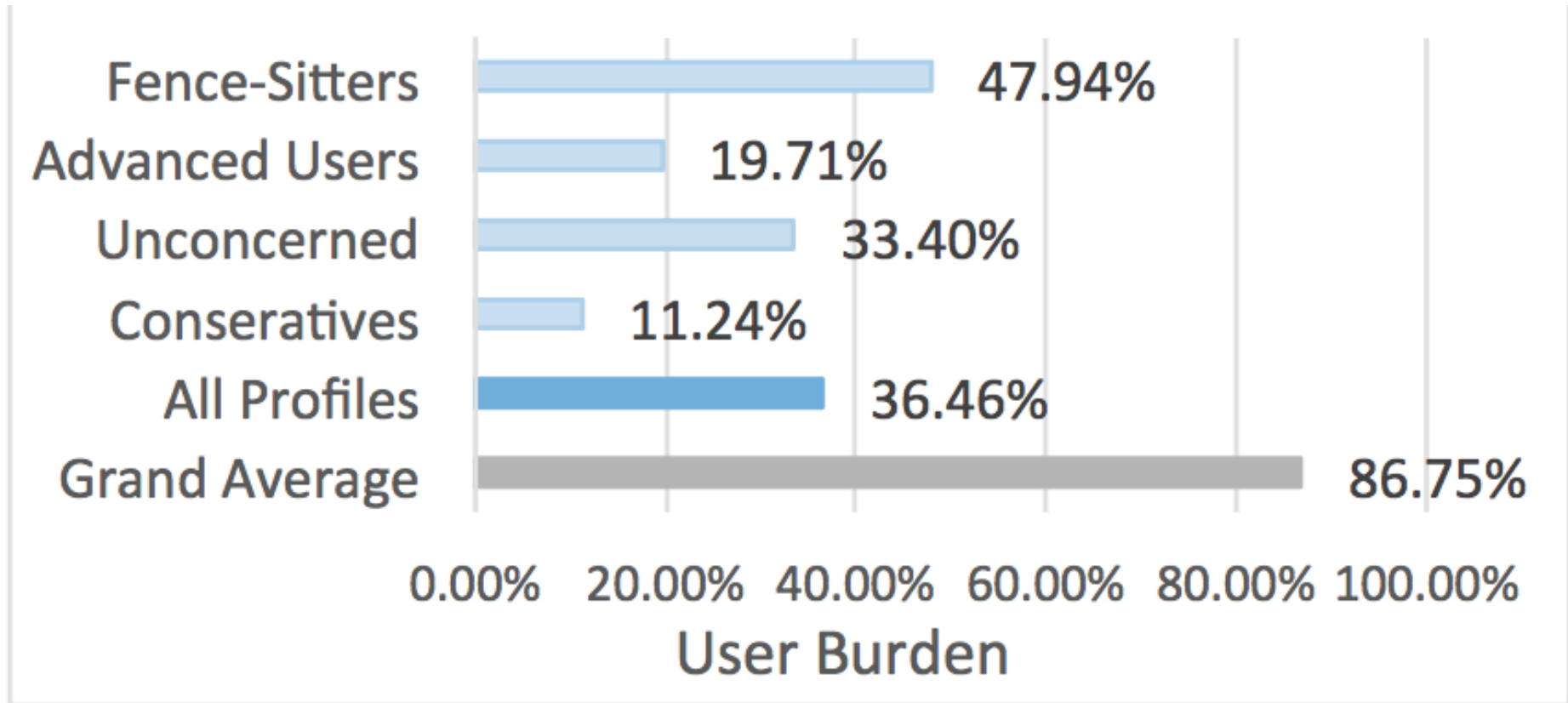
Potential Benefits - Accuracy



“Grand Average”: Results obtained with “one-size-fits-all” profile

Potential Benefits – User Burden

Only Prompt the User When the Prediction is Uncertain



“Grand Average”: Results obtained with “one-size-fits-all” profile

LBE Privacy Guard

- ❑ Fine-grained control of Android Permissions.
- ❑ Available on rooted Android phones.
- ❑ Managing 12 Permissions:
 - "Send SMS", "Phone Call", "Phone State", "Call Monitoring", "SMS DB", "Contact", "Call Logs", "Positioning", "Phone ID", "3G Network", "Wi-Fi Network" and "ROOT".
- ❑ Settings that users can choose:
 - "Allow", "Deny", "Ask"

Emergency calls only ... 0.01K/s 1:29PM

< Manage App Permissions

Enable Permission manager ☒

Monitoring 32 apps

SECURITY

Send SMS
7 apps are allowed to send SMS

Phone Call
5 apps are allowed to make, answer or hang up phone calls

Phone State
6 apps are allowed to get phone state

Call Monitoring
8 apps are allowed to monitor incoming, outgoing calls and control ringer volume.

PRIVACY

SMS
3 apps are allowed to access your SMS inbox

Contacts
8 apps are allowed to access your contact list

Call Log

Emergency calls only ... 0.11K/s 1:15PM

Permissions Apps Logs

Monitoring 21 apps

 Angry Birds
2 permissions


 MiTalk
8 permissions Trusted

 QQ
8 permissions


 Weibo
8 permissions

 Voice assist
8 permissions Trusted

 WeChat
6 permissions Trusted

 Google Play Store
3 permissions

 Douban Movie
2 permissions

 SoundHound
4 permissions

Emergency calls only ... 0.03K/s 1:19PM

< Google Play Store version 4.1.10


I trust this App ☐

MIUI will monitor this app

Autostart permission ☐


Allow autostart operation

SECURITY RELATED


Send SMS 

Send SMS directly

PRIVACY RELATED

Positioning 

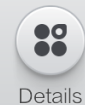
Locale your phone through network or GPS

Phone ID 

Get your phone ID, including IMEI, IMSI, etc

LOGS

obtain IMSI number Allow



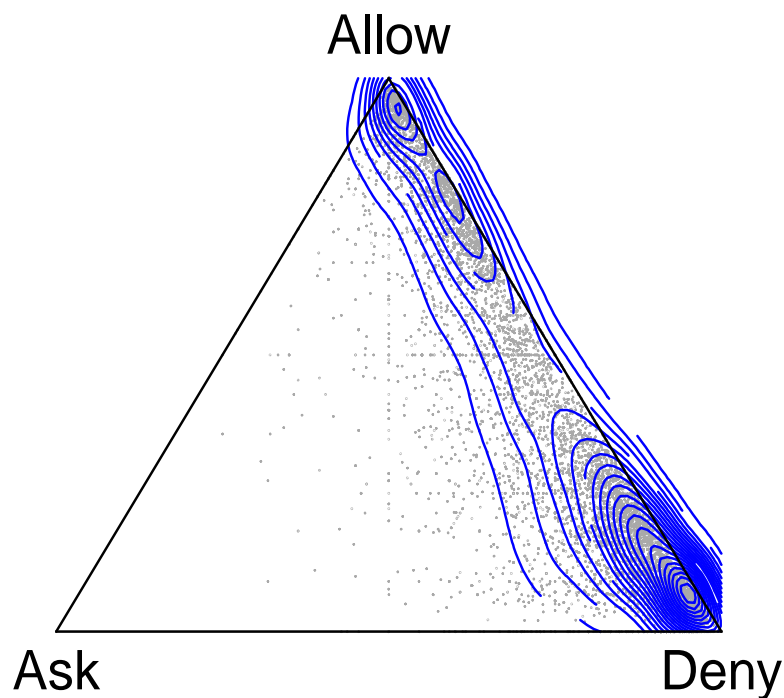
Details

Predicting settings for each user

- $f: \{user, app, permission\} \rightarrow decision$
- Predicting “Allow” and “Deny”.
- Large-scale corpus:
 - **239K users, 12K apps, 14.5M records**

Diversity of users' preferences

- ❑ 22.66 apps per user
- ❑ 3.19 common apps per pair of users
- ❑ 3.03 permissions per app
- ❑ Agreement of users' decisions:
 - 63.9% of app-permission pairs have 80% agreement.
(If we consider pairs with ≥ 5 users, the percentage drops to 51.4%)

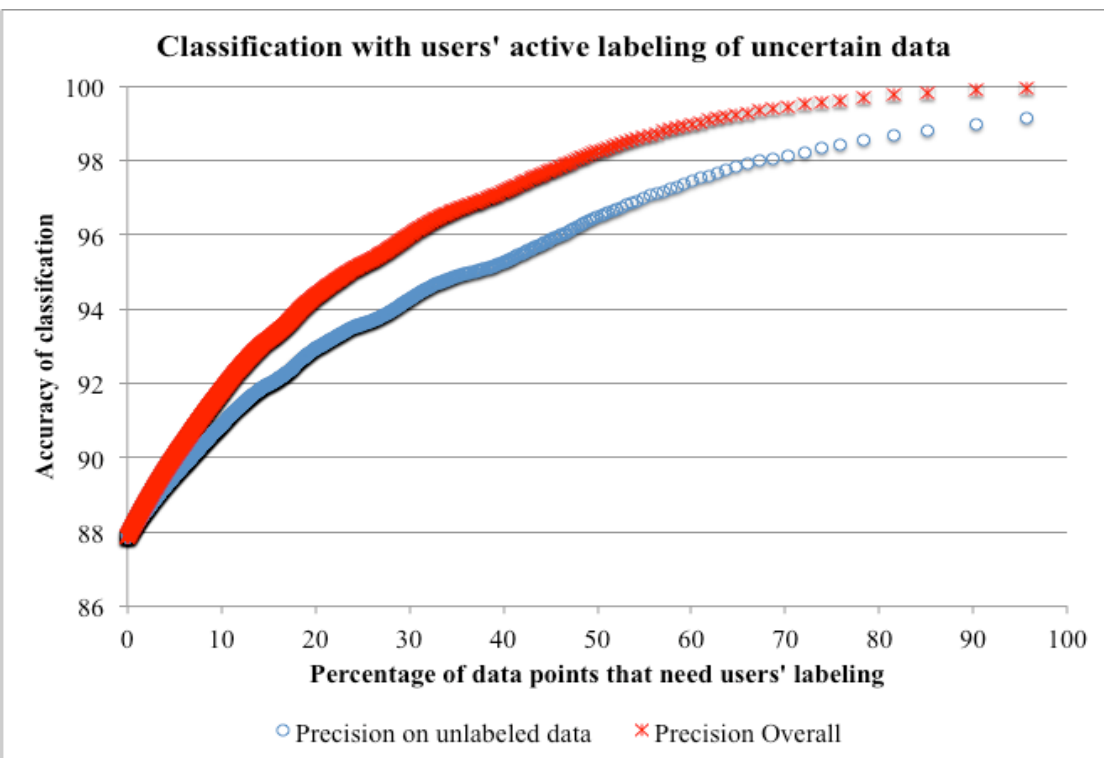


Each dot is an app-permission pair with its Ask/Deny/Allow mix of users

Predicting settings for each user

- $f: \{user, app, permission\} \rightarrow decision$
- Predicting “Allow” and “Deny”.
- Large-scale corpus:
 - **239K users, 12K apps, 14.5M records**
- Split of training & testing on users
 - Training set: users’ decisions are all known
 - Testing set: users’ decisions of only 20% of their apps are known
 - 10-fold cross validation
- Linear-kernel SVM
 - L2-loss dual SVM classification

Pure Prediction vs. Interactive Model



With more labeling of users, we can increase the accuracy of our predictions.

If users can label an additional 10% of their permission decisions, the **prediction accuracy will climb from 87.8% to 91.8%...and that's only 6 questions...**

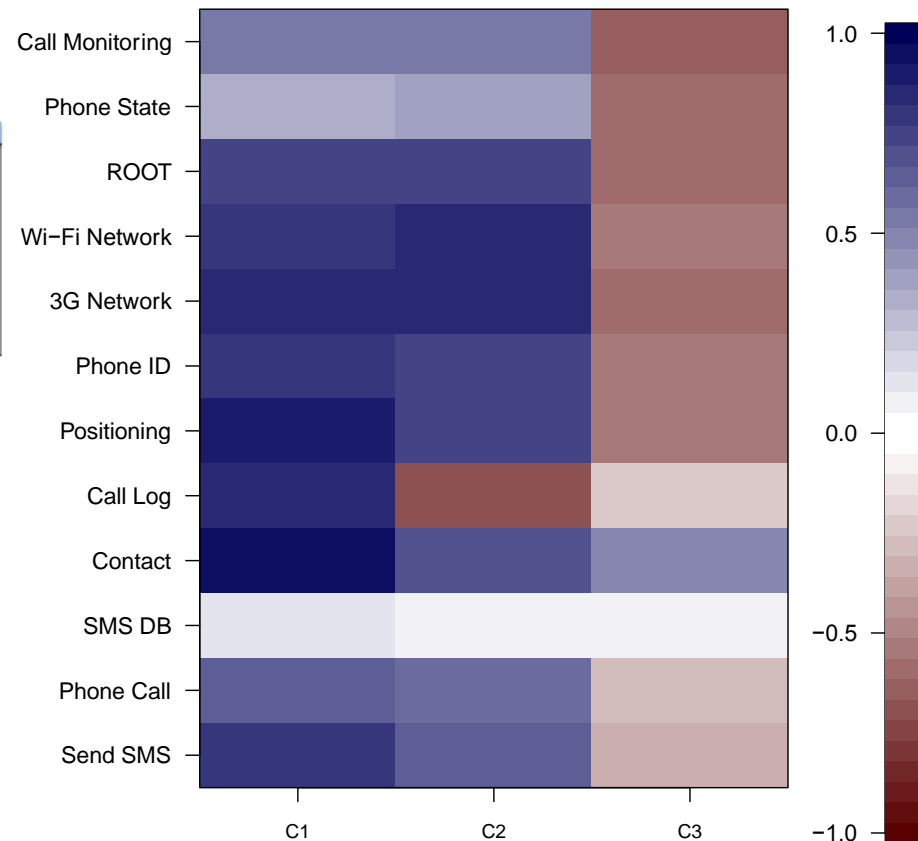
At 20% (about 12 questions), accuracy climbs to 94%!

And Profiles Can Help too (e.g. K=3)

Profile 1 (24.4%)		
✓ Call Log + ✓ 3G / Wi-Fi		86%
✓ Call Log + ✓ ROOT		84%
✓ Call Log + ✓ Positioning		84%
✓ Call Log + ✓ Phone ID		83%
✓ Call Log + ✓ Phone State		82%

Profile 2 (44.0%)		
✗ Call Log + ✓ 3G / Wi-Fi		80%
✗ Call Log + ✓ ROOT		77%
✗ Call Log + ✓ Phone State		76%
✗ Call Log + ✓ Call Monitoring		75%
✗ Call Log + ✓ Positioning		75%

Profile 3 (31.6%)		
✗ Wi-Fi / 3G Network		69%
✓ SMS DB + ✗ Wi-Fi / 3G		68%
✗ Phone ID		66%
✗ ROOT Privileges		65%
✗ Phone ID + ✓ SMS DB		82%



Profile Descriptions & Heat map of users' average decisions
when K=3

Concluding Remarks - I

- ❑ Mobile App Privacy is critical to the reputation of app stores
- ❑ Growing number of APIs and complex data flows
- ❑ **Finer privacy settings are overwhelming**
 - And the current ones also ignore “purpose” which critically impacts people’s preferences
- ❑ Preferences based on **app-permission-purpose**, while **seemingly more complex**, can help develop **deeper preference models** and ultimately **simplify user decisions**

Concluding Remarks - II

Long-term goal: **Intelligent Privacy Assistants**

- Help scale to interactions with a large number of apps and services
- Learn user preferences
- Can selectively enter in dialogues with users and **nudge them towards safer practice**

Did you know that over the past 24 hours your apps **shared your location 37 times with 5 different Profiling Companies?**

Concluding Remarks - III

- ❑ The **Internet of Things** will make the need for better privacy technologies even more critical
- ❑ Towards more practical laws and regulations
 - Today's "Notice and Choice" framework used in the US does not work – **no one reads privacy policies**
 - Two guys in a garage can't be expected to articulate a privacy policy
 - **App stores & SDK providers have an important role to play here**

Q&A



<http://mcom.cs.cmu.edu>

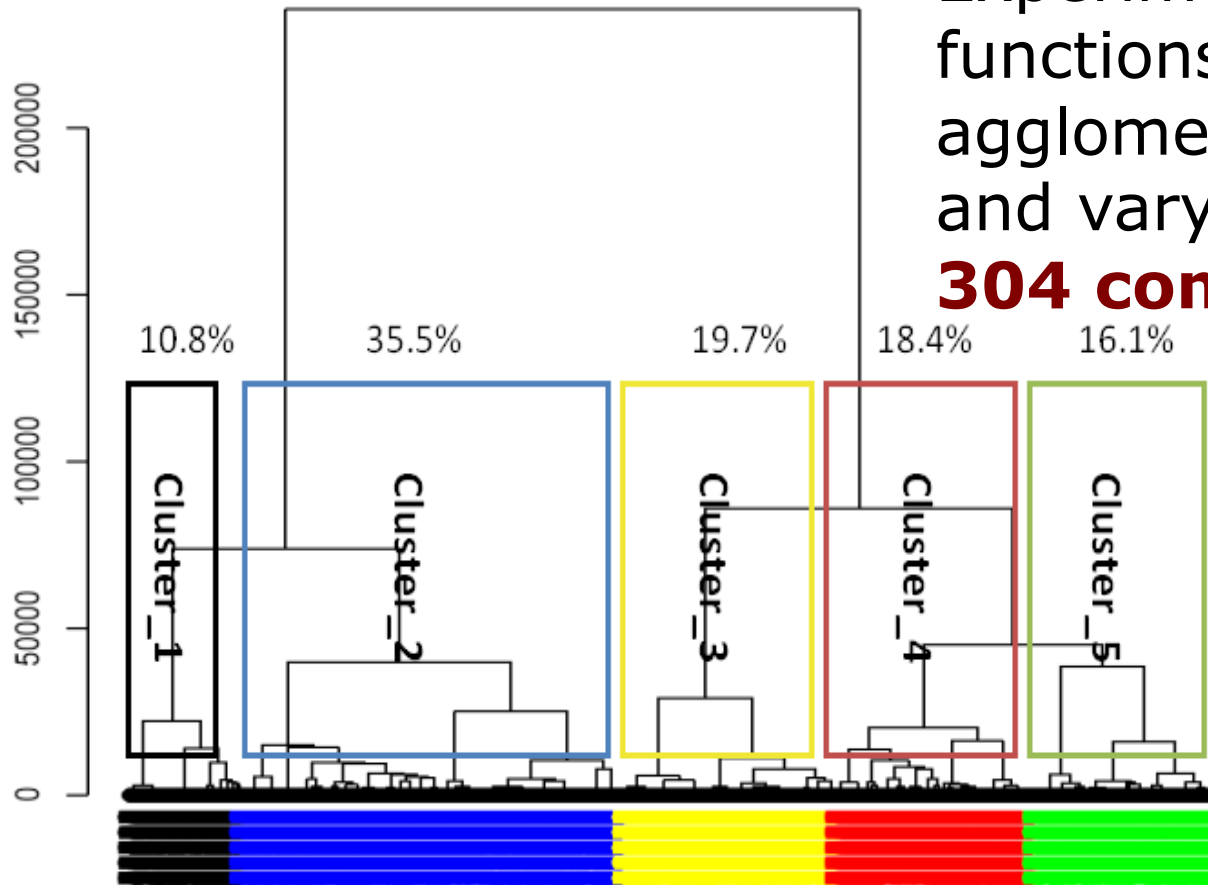


Additional Slides

Identify Patterns in App Behaviors

- **Insight:** Privacy decisions reflect subjective tradeoffs users make between utility and privacy
- **A first study:** Looking for common patterns in usage of sensitive resources
 - Taking into account purpose information

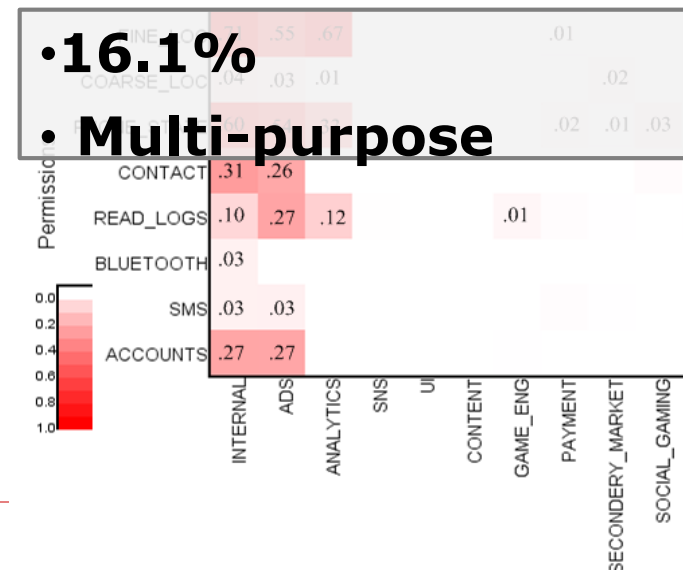
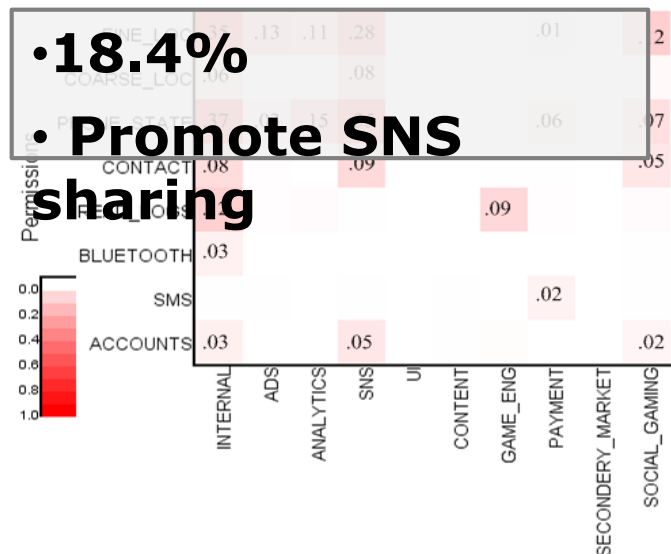
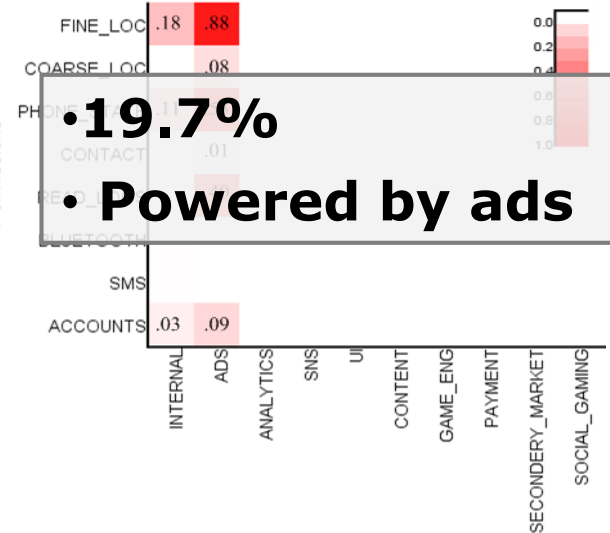
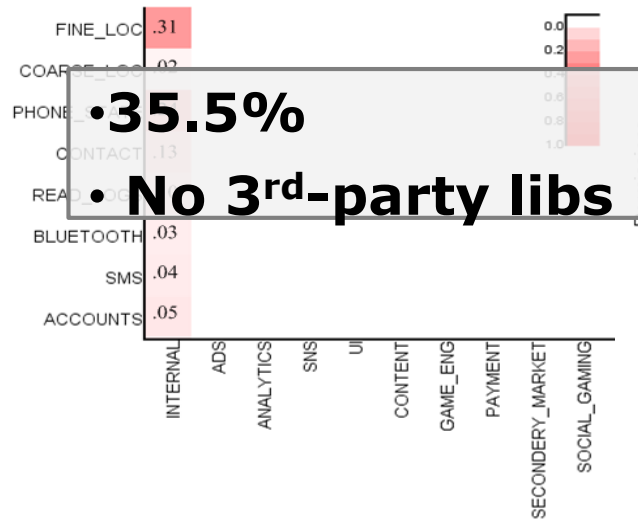
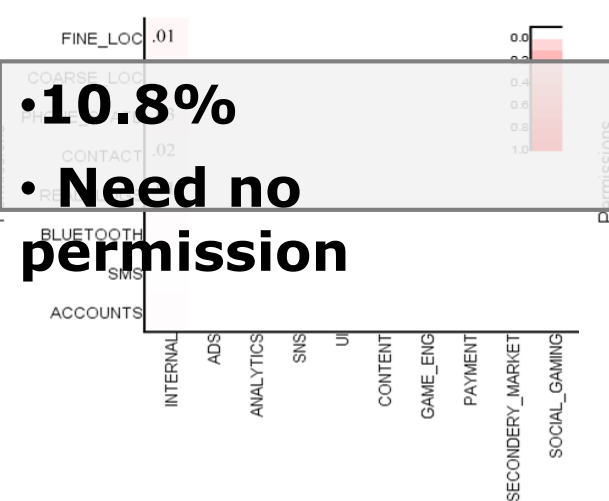
Hierarchical Clustering of Apps



Experiment 4 distance functions, 4 agglomerative methods, and vary k from 2-20 → **304 combinations**

Dendrogram of hierarchical clustering with **Canberra** distance and **Ward's** method when **k=5**

Characteristics of Each Cluster



Characteristics of Each Cluster

• **10.8%**

• **Need no permission**

• **35.5%**

• **No 3rd-party libs**

• **19.7%**

• **Powered by ad**

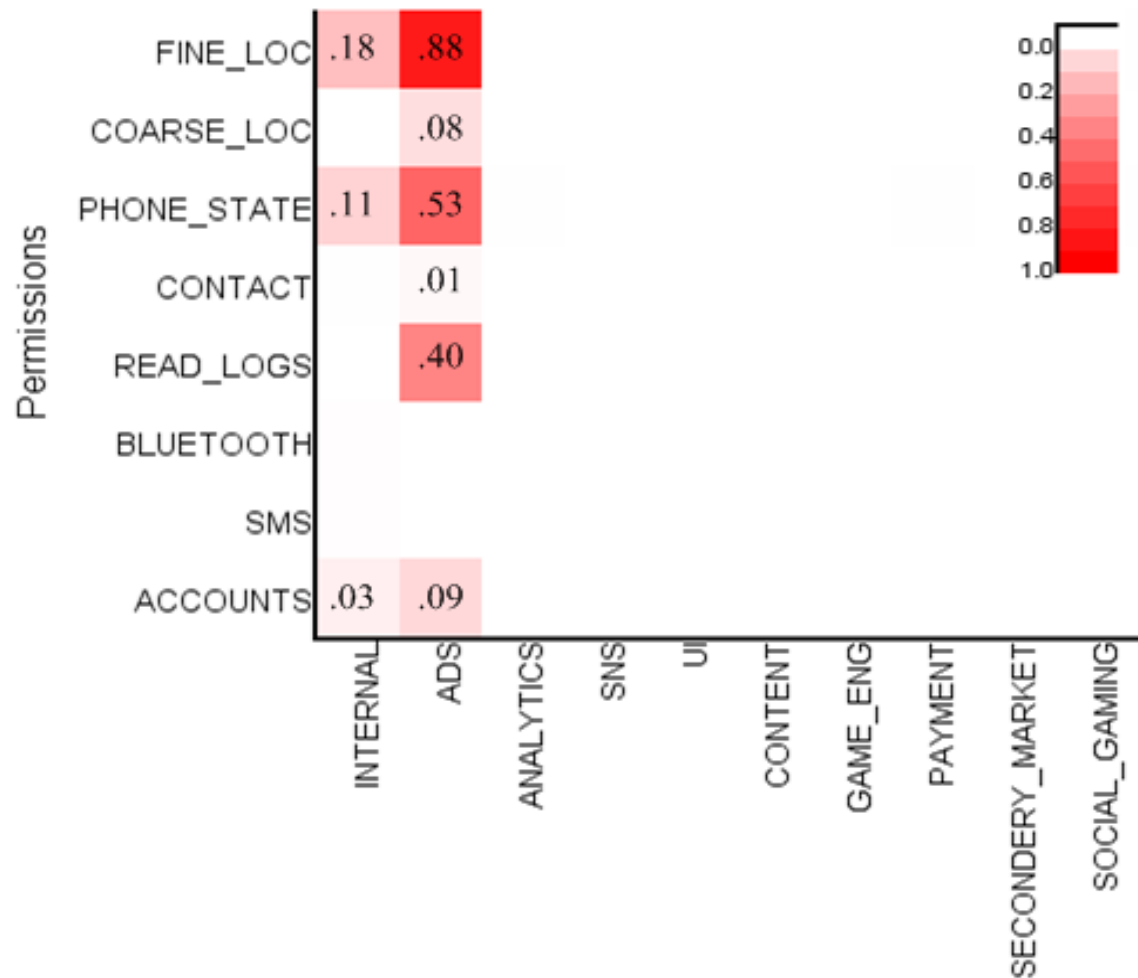
• **18.4%**

• **Promote SNS sharing**

• **16.1%**

• **Multi-purpose**

Cluster 3 – Ad powered (19.70%)



Mouse, HD
wallpapers Live,
Tunee Music

- Sensitive resources used mainly for delivering ads
- Privacy risks: information aggregation by ad agencies

Characteristics of Each Cluster

• **10.8%**

• **Need no permission**

• **35.5%**

• **No 3rd-party libs**

• **19.7%**

• **Powered by ad**

• **18.4%**

• **Promote SNS sharing**

• **16.1%**

• **Multi-purpose**

Next Step

- Looking for other clusters that capture other aspects of the utility users derive from different categories of apps
 - e.g. also differentiating between games, productivity tools, etc.
- Deriving user profiles based on these categories