Are You Being Tracked? What Can you Do?

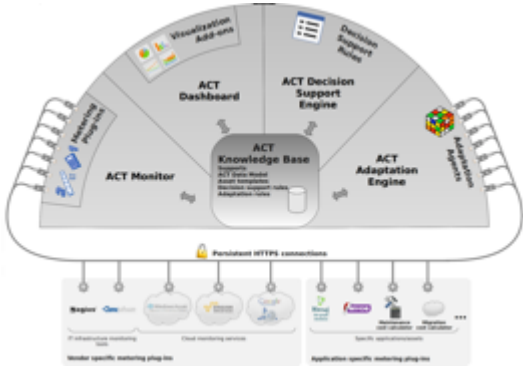
Aruna Seneviratne &
Suranga Seneviratne+++

Australian Government
Department of Broadband, Communications and the Digital Economy
Australian Research Council

NICTA Funding and Supporting Members and Partners

NICTA in Brief

- Australia's National Centre of Excellence in Information and Communication Technology
- Five Research Labs:
  - ATP: Australian Technology Park, Sydney
  - NRL: UNSW, Sydney
  - CRL: Canberra
  - VRL: Melbourne
  - QRL: Brisbane
- 700 staff including 300 PhD students
- Budget: ~$90m/y from Fed/State Gov and industry



Motivation

- The personal information collected from the sensors, and use of mobile devices
  - Provision of personalised services to the users
- Personalisation comes at a cost to user's security and privacy

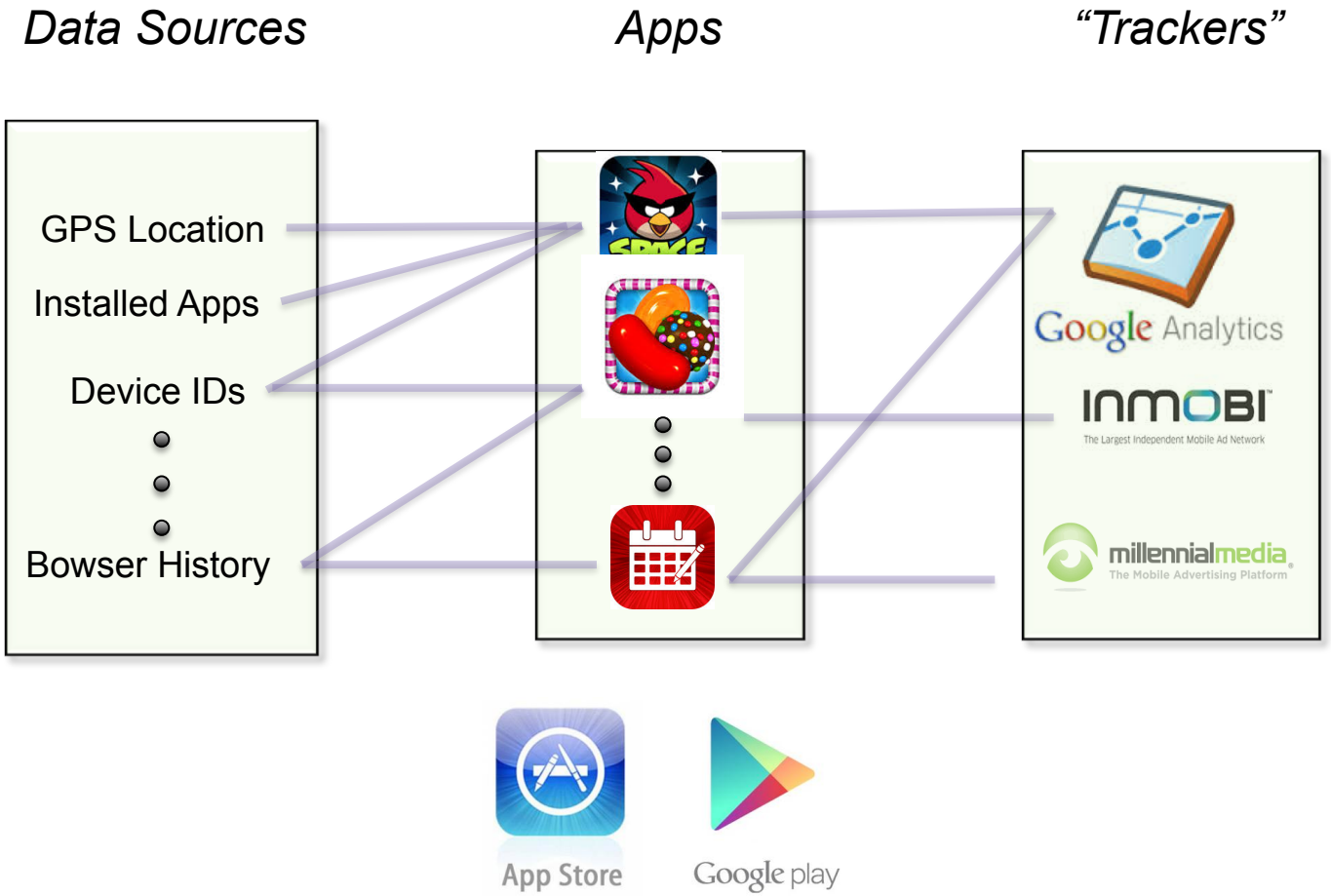
Challenge

- How to safe guard the security and privacy of the users, whilst still providing the full benefits of personalized services
  1. Provide information to *users to make them informed decisions* :utility vs. loss of security/privacy
  2. Have tools to detect fraudulent apps
  3. Methods of extracting information whilst guaranteeing security and privacy: privacy preserving analytics

Challenge

- How to safe guard the security and privacy of the users, whilst still providing the full benefits of personalized services
  1. Provide information to *users to make them informed decisions* :utility vs. loss of security/privacy
  2. Have tools to detect fraudulent apps
  3. Methods of extracting information whilst guaranteeing security and privacy: privacy preserving analytics

Today: Users in the "Dark"



Example - #1

- It is possible to identify user traits very easily
- A single snapshot of apps installed on a smartphone!
  - Apptronomy
    - Upon installation, lists and uploads the user installed apps to a server
    - Generates a random ID for that installation instance
  - Group of volunteers and users through Amazon Mechanical Turk
    - User traits through a brief questionnaire
  - Crawled two popular social app discovery sites: *Appbrain* and *Appaware*

	Appbrain	Appaware	Apptronomy
# of users	8653	841	369
# of apps	85770	24254	6341
# of installations	705004	94024	15710
Average # of apps/user	81	112	43
Median # of apps/user	51	75	34

Example - #1.1

- Trained SVM classifiers
  - app description as the input and predict whether the given app is relevant to that particular trait

	Precision			Recall		
	>0	> 1	> 2	>0	> 1	> 2
Language	62%	86%	82%	33%	25%	19%
Country						
Top-25	97%	100%	100%	17%	8%	5%
Top-50	98%	96%	94%	29%	12%	7%
Top-75	40%	63%	68%	37%	15%	9%
Religion	90%	100%	100%	24%	5%	3%
Is single?	70%	100%	100%	26%	10%	2%
Is a parent?	53%	78%	100%	26%	10%	7%

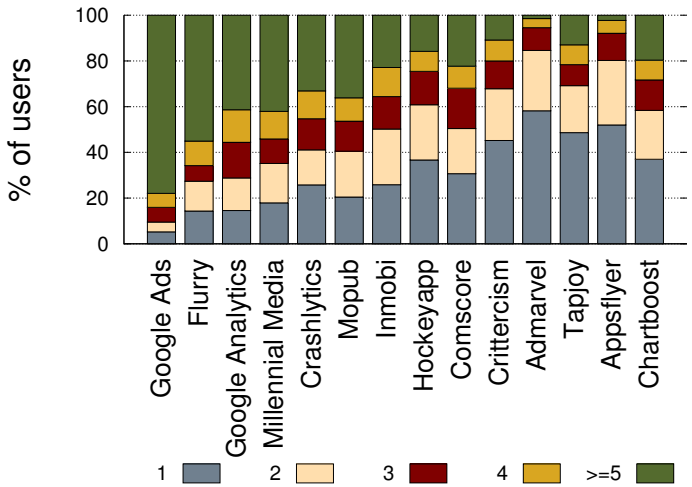
- Installed apps in smartphones **can** infer user traits

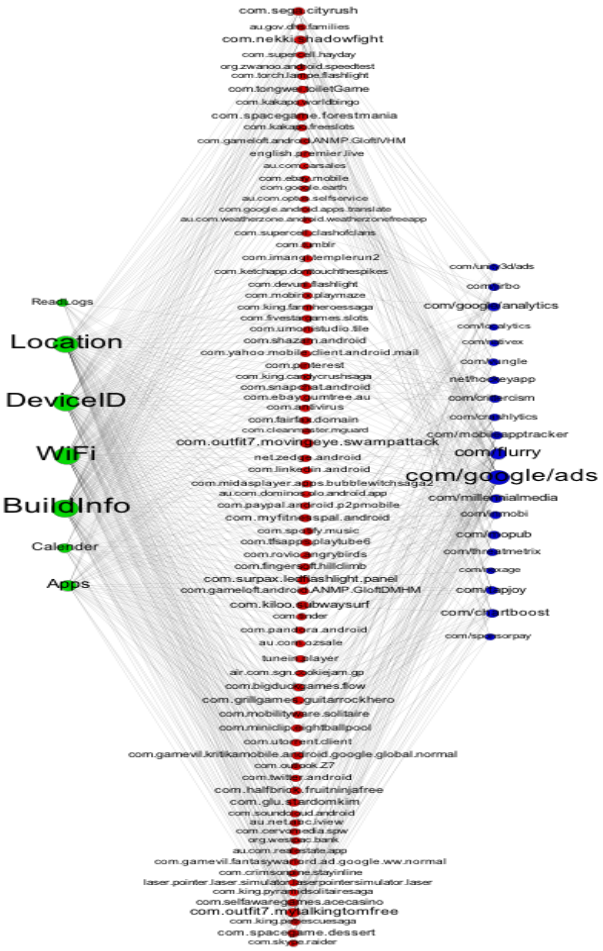
S. Seneviratne, A. Seneviratne,. Mahanti, P. Mohapatra. "Your Apps Are What You Are: User Traits Through Installed Smartphone Apps" ACM SIGMOBILE Mobile Computing and Communications Review 18 (3), 55-61, 2015.

Example #2

- A few know a lot
  - Identified the top-100 free and paid apps from four countries representing four geographical regions
    - 275 unique free and 234 unique paid apps
  - For all the apps found in users' app downloaded the APK files - 3,605
  - Two analysis tools to identify the embedded trackers and the API calls executed by the trackers
  - Permissions are abstract and may not necessarily represent the full implications

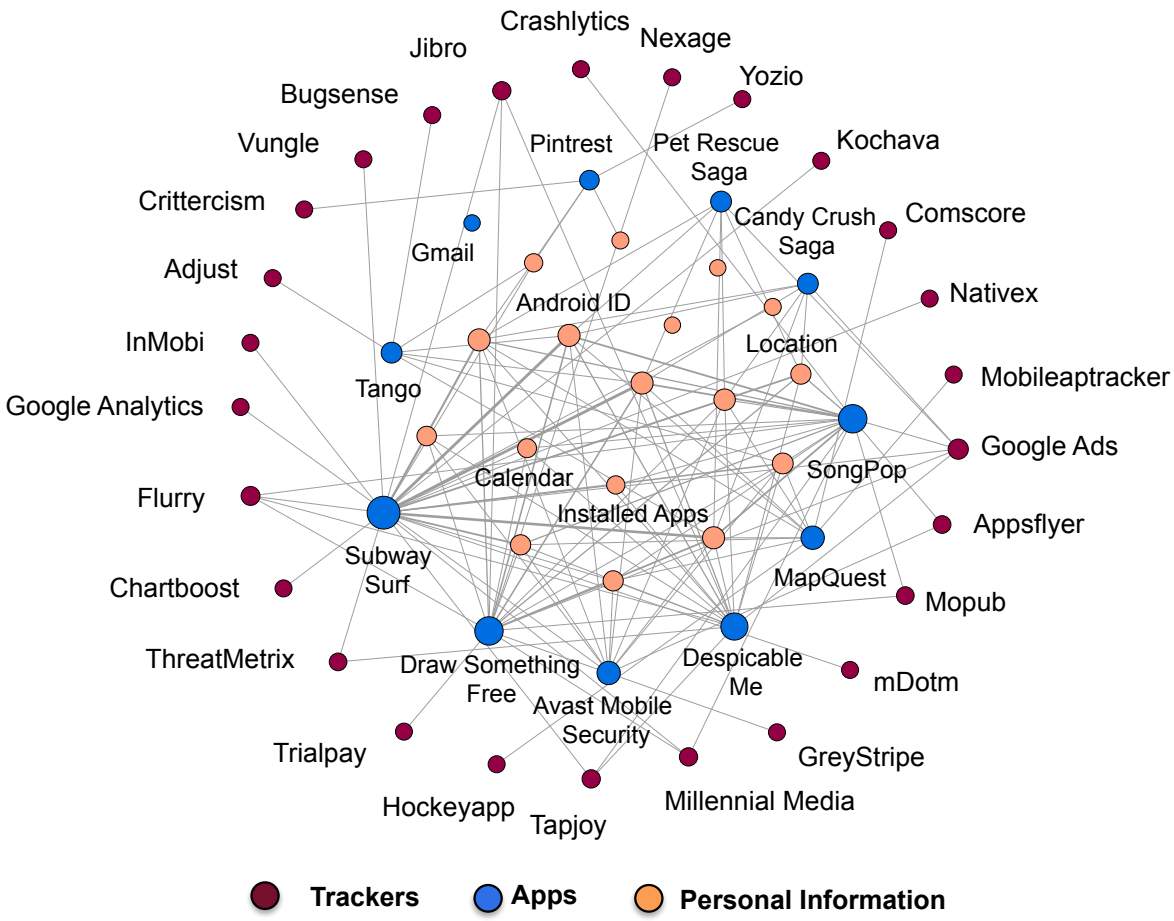
Example #2.1





S. Seneviratne,H. Kolumunna A. Seneviratne,"A Measurement Study on Tracking in Paid Mobile Applications" NICTA Technical Report 2015-8, February, 2015

Single User



11 apps exposed 26 trackers !!

Android Malware Removed From Google Play Store After Millions of Downloads

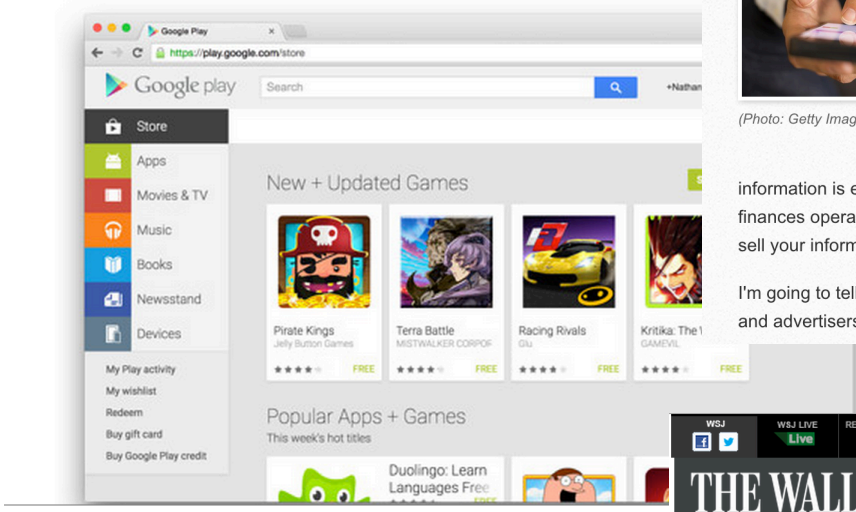
NICTA

ARTICLE | COMMENTS (9)

ANDROID AVAST GOOGLE GOOGLE PLAY MALWARE

Email Print 33 62

By NATHAN OLIVAREZ-GILES CONNECT



What cellphone companies don't want you to know

Kim Komando, Special for USA TODAY *7:04 a.m. EDT March 13, 2015*

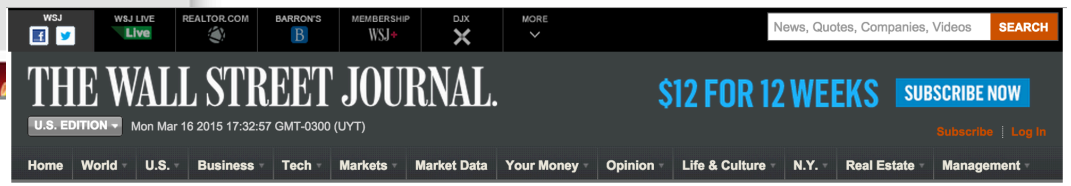
(Photo: Getty Images/Thinkstock)

16877 CONNECT | 547 TWEET | 171 LINKEDIN | 22 COMMENT | EMAIL | MORE

Recently, AT&T surprised everyone when it added a new option to its GigaPower fiber Internet service: privacy. Yes, for just $29 more a month AT&T promises it WON'T sell your search and browsing history to advertisers. How generous.

While there's still some doubt about how private your information is even after you pay the $29, at least AT&T is being honest about how it finances operations. The truth is, the major cellphone carriers are more than happy to sell your information to advertisers and serve you targeted ads over their networks.

I'm going to tell you how to stop them, and at the end I'll discuss other ways carriers and advertisers are working to get your information.



What They Know - Mobile

+ More

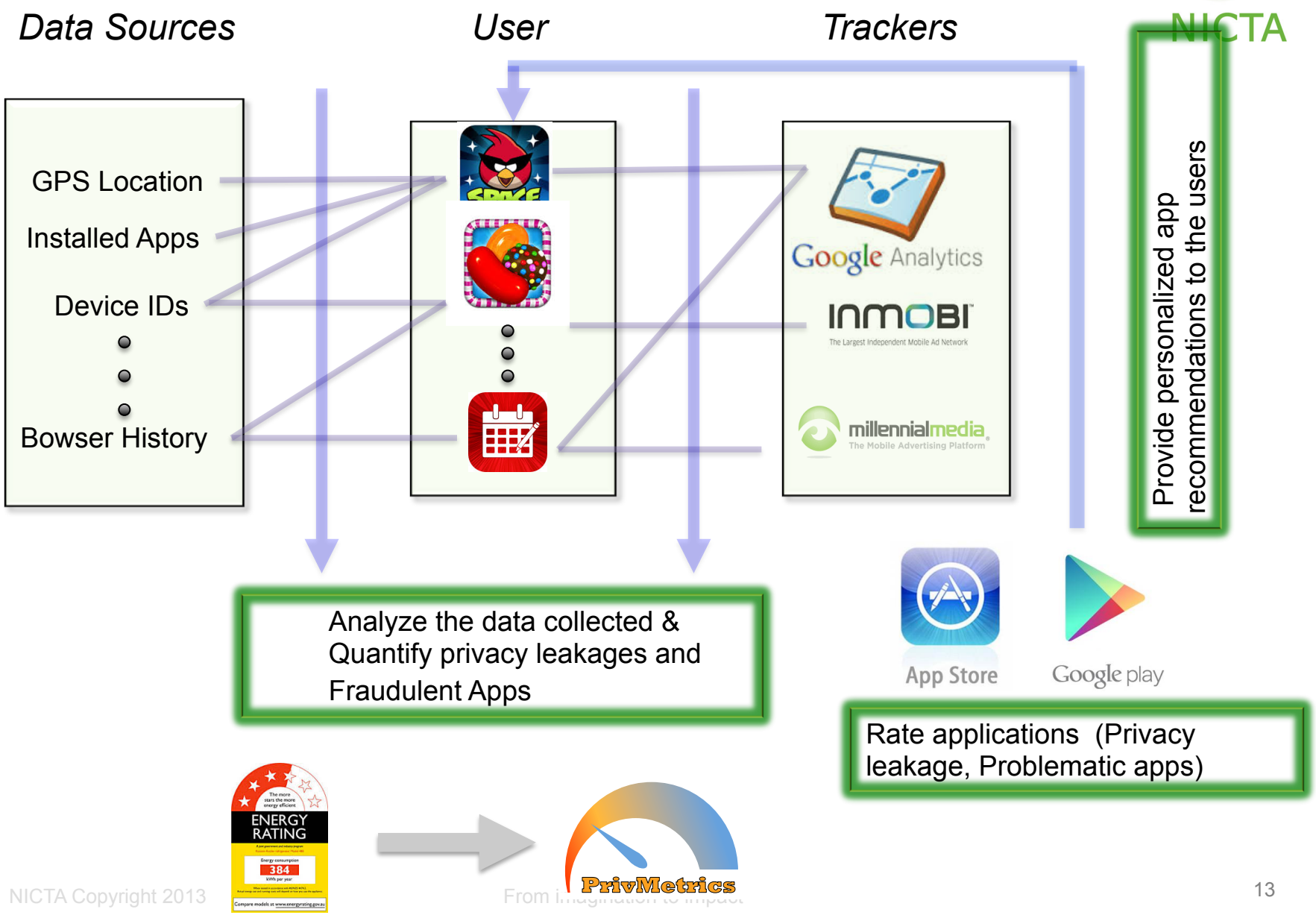
Marketers are tracking smartphone users through "apps" - games and other software on their phones. Some apps collect information including location, unique serial-number-like identifiers for the phone, and personal details such as age and sex. Apps routinely send the information to marketing companies that use it to compile dossiers on phone users. As part of the What They Know investigative series into data privacy, the Journal analyzed the data collected and shared by 101 popular apps on iPhone and Android phones (including the Journal's own iPhone app). This interactive database shows the behavior of these apps, and describes what each app told users about the information it gathered.

Recent Stories

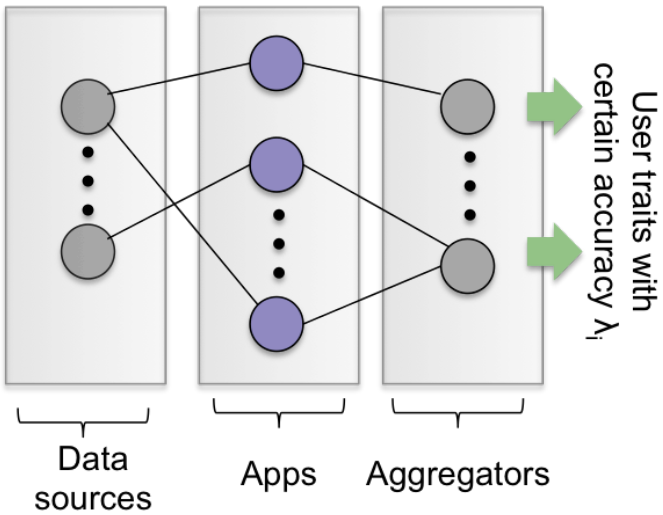
More views of the data »

APPS KIDS THE TOP 50 SITES

Basic Idea: Informed Decisions



Rating of Apps



For aggregator *i*, let

$\Lambda^T = (\lambda_1\ \lambda_2\ \ldots \lambda_p)$, the accuracy vector for user trait *p* and
$U^T = (u_1\ u_2\ \ldots u_p)$, the vector representing users willingness to share trait *p*

"*Privacy level*" w.r.t aggregator *i*,

$$X_i = f(\Lambda^T, U^T)$$

Then "Overall *Privacy Level*"

$$P = g(X_1 \ldots X_D)$$

Where *D* is the number of aggregators and *g* is the weighted mean function.

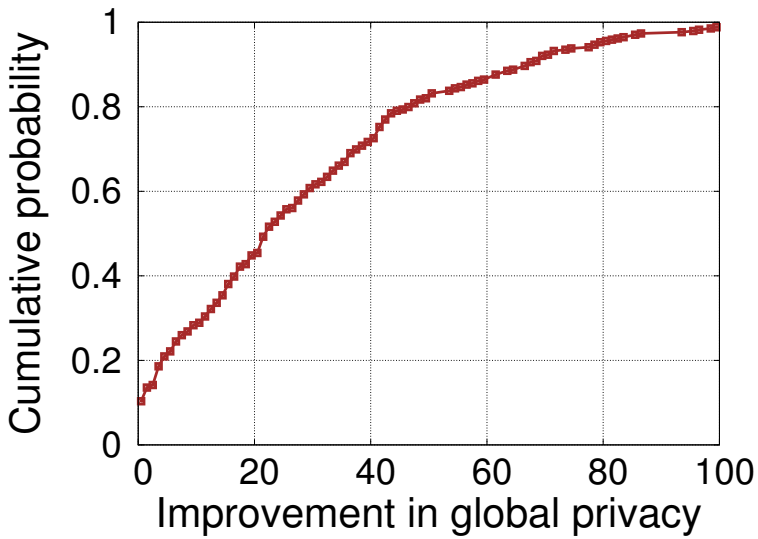
Objective: Maximize *P*, subject to

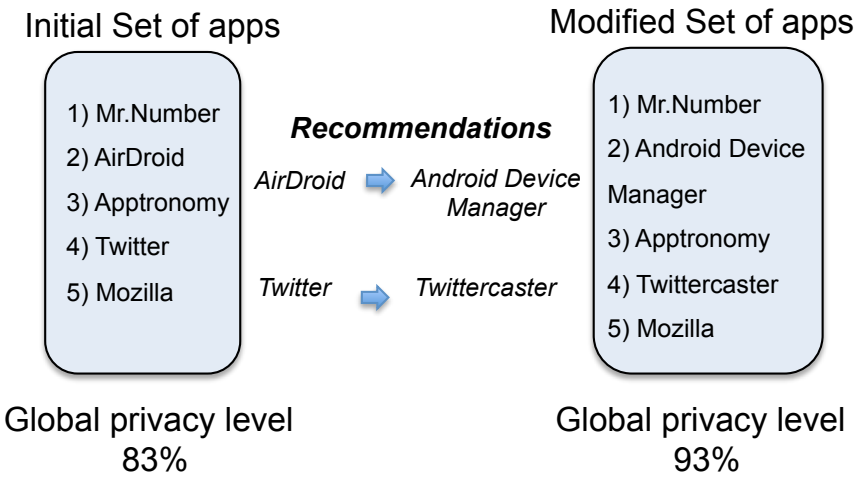
A_i in $\{A_{i,I}\ ;\ A_{c1,i}\ ;\ \ldots\ldots\ ;\ A_{cj,i}\}$, $i=1:K$
where

- *K* is the number of apps
- $A_{i,i}$ is the original application and $A_{cj,i}$ s are the apps providing a similar function to the original.

Solved using "Steepest Ascent Hill Climbing"

Recommendation of Apps





50% of the users gained over 20% increment in "*overall privacy level*"

Example recommendation by PrivMetrics

Challenge

- How to safe guard the security and privacy of the users, whilst still providing the full benefits of personalized services
  1. Provide information to *users to make them informed decisions* :utility vs. loss of security/privacy
  2. Have tools to detect fraudulent apps
  3. Methods of extracting information whilst guaranteeing security and privacy: privacy preserving analytics

Detecting Fraudulent Apps (1)

NICTA

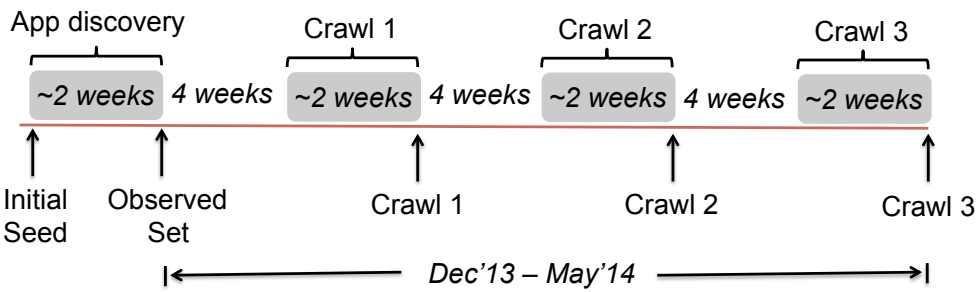
- State-of-the-art mobile malware detection is only reactive!
  - *(based on Known malware DBs, Signature Comparison, User feedback)*
- Early detection can reduce further damage
- Challenges
  - *Limited amount of data (No user reviews or ratings)*
  - *Predictions need to be precise (Legitimate apps must not be penalized)*
  - *Ability to quickly analyze a large number of apps (Fast approval for developers)*

Angry Birds

Angry Purrs!

Detecting Fraudulent Apps (2)

- Discover
  - Functionally similar apps
  - Other apps by the same developer
- Metadata such as app name, app description, and app category for all the apps



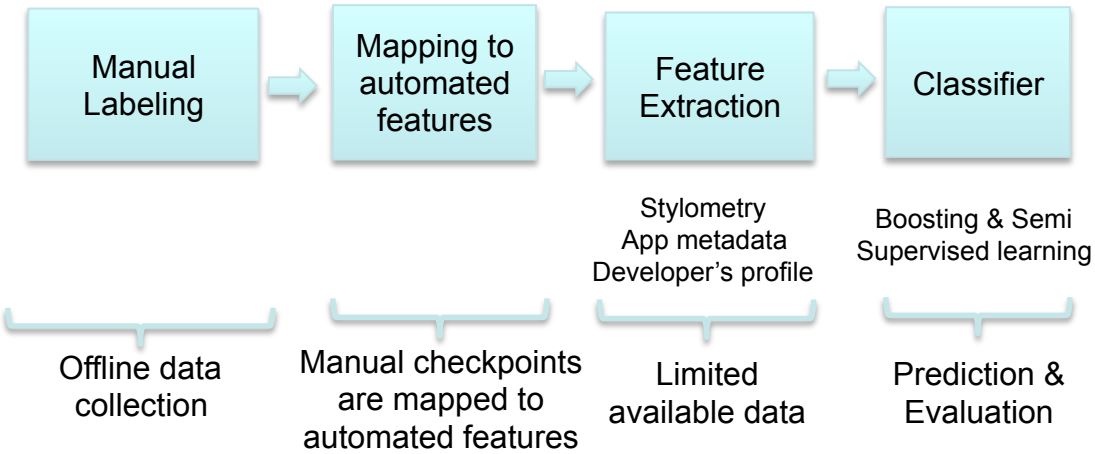
Set	Number of apps
Observed set ($\mathbb{O}$)	232,906
Crawl 1 ($\mathbb{C}_1$)	6,566
Crawl 2 ($\mathbb{C}_2$)	9,184
Crawl 3 ($\mathbb{C}_3$)	18,897

Detecting Fraudulent Apps (3)

- Identified the reasons for app removal
  - Consulting numerous market reports
  - Examining the policies of the major app markets

Reason	Description
Spam	These apps often have characteristics such as unrelated description, keyword misuse, and multiple instances of the same app. Section 4 presents details on spam app characteristics.
Unofficial content	Apps that provide unofficial interfaces to popular websites or services (E.g., an app providing an interface to a popular online shopping site without any official affiliation).
Copyrighted content	Apps illegally distributing copyrighted content.
Adult content	Apps with explicit sexual content.
Problematic content	Apps with illegal or problematic content. E.g., Hate speech and drug related.
Android counterfeit	Apps pretending to be another popular app in the Google Play Store.
Other counterfeit	A counterfeit app, for which the original app comes from a different source than Google Play Store (E.g., Apple App Store)
Developer deleted	Apps that were removed by the developer.
Developer banned	Developer's other apps were removed due to various reasons and Google decides to ban the developer. Thus all of his apps get removed.

Detecting Fraudulent Apps (4)



- Aggressive classifier
  - ~70% of the removed apps and 55% of the other apps to be spam
- conservative classifier
  - 6% to12% of the removed apps 2.7% of the other apps a

S. Seneviratne, A. Seneviratne, M.A. Kaafar, A. Mahanti, P. Mohapatra. "Early Detection of Spam Mobile Apps". To appear in ***ACM WWW'15****.*

Conclusions

NICTA

- With personal information collected from the sensors, and use of mobile devices its obvious that more protection is needed
- We believe that this is bets done with the user in the centre of the decision making process
- Eneregy “Star Rating” scheme for electrical good
- *Privmetrics* provides a framework for developing such a rating system which can be extended to provide other services as well