

Differentiating Smartphone Users by App Usage

P. Welke, I. Andone, K. Błaszkiwicz, A. Markowetz

UbiComp '16

Introduction

Differentiating Smartphone Users by App Usage

- Most likely, you don't want to be spied upon, right?
 - you can disable lots of things on your phone
 - you can add a firewall
 - you can spoof your hardware identifiers

Introduction

Differentiating Smartphone Users by App Usage

- Most likely, you don't want to be spied upon, right?
 - you can disable lots of things on your phone
 - you can add a firewall
 - you can spoof your hardware identifiers
- But would you be willing to *change your behavior*?

How can we Measure Differences in Behavior?

Differentiating Smartphone Users by App Usage

- You use a certain set of apps, lets call that your *app signature*
 - can be leaked by apps on your phone
 - can be inferred from network traffic

How can we Measure Differences in Behavior?

Differentiating Smartphone Users by App Usage

- You use a certain set of apps, lets call that your *app signature*
 - can be leaked by apps on your phone
 - can be inferred from network traffic
- Let's call you *anonymous*, if there is another user with the same app signature as you

How can we Measure Differences in Behavior?

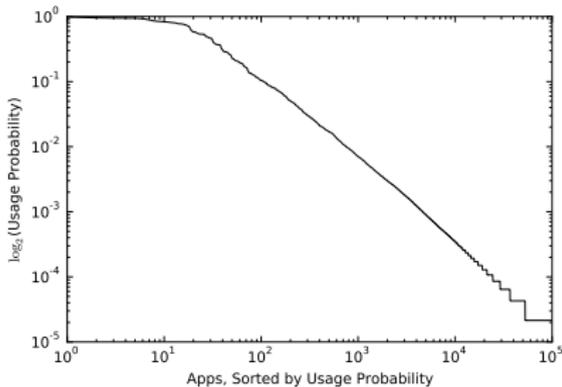
Differentiating Smartphone Users by App Usage

- You use a certain set of apps, lets call that your *app signature*
 - can be leaked by apps on your phone
 - can be inferred from network traffic
- Let's call you *anonymous*, if there is another user with the same app signature as you
- We will also look at the Hamming distance between two app signatures
 - Measures the number of apps that are used on exactly one of the two phones
 - If there is a large distance between two signatures, then the behavior of the users is different

App Popularity on Android

Differentiating Smartphone Users by App Usage

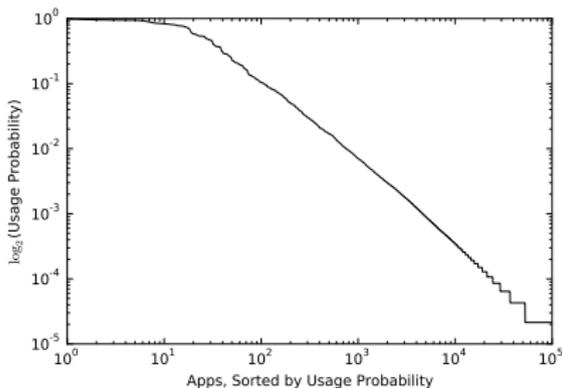
- There are *lots* of apps. Most are extremely unpopular.



App Popularity on Android

Differentiating Smartphone Users by App Usage

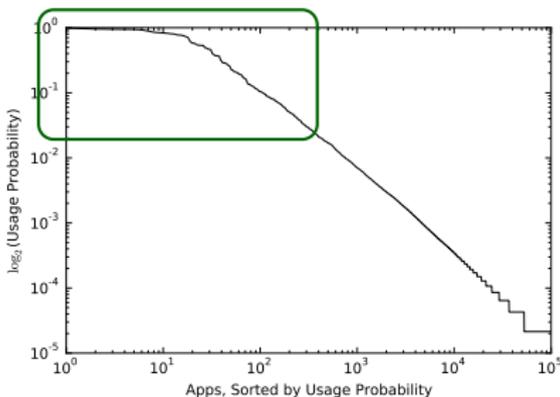
- There are *lots* of apps. Most are extremely unpopular.
- You most likely use at least one unpopular app.



App Popularity on Android

Differentiating Smartphone Users by App Usage

- There are *lots* of apps. Most are extremely unpopular.



- You most likely use at least one unpopular app.
- Ok, maybe you can cope with using only more popular apps, *let's say the top 500...*

The Dataset

Differentiating Smartphone Users by App Usage

- app usage data of 46 726 users

The Dataset

Differentiating Smartphone Users by App Usage

- app usage data of 46 726 users
- observation periods from 1 to 216 days

The Dataset

Differentiating Smartphone Users by App Usage

- app usage data of 46 726 users
- observation periods from 1 to 216 days
- average observation period 49 days

The Dataset

Differentiating Smartphone Users by App Usage

- app usage data of 46 726 users
- observation periods from 1 to 216 days
- average observation period 49 days
- part of a larger dataset (www.mental.org)

The Dataset

Differentiating Smartphone Users by App Usage

- app usage data of 46 726 users
- observation periods from 1 to 216 days
- average observation period 49 days
- part of a larger dataset (www.mental.org)

age	# users
$0 \leq x < 12$	292
$12 \leq x < 17$	10283
$17 \leq x < 21$	10970
$21 \leq x < 25$	7398
$25 \leq x < 30$	6560
$30 \leq x < 35$	4105
$35 \leq x < 40$	2531
$40 \leq x < 50$	3244
$50 \leq x < 70$	1300
$70 \leq x < 100$	43

Top 500 Apps

Differentiating Smartphone Users by App Usage

- How many users are anonymous?

Top 500 Apps

Differentiating Smartphone Users by App Usage

- How many users are anonymous?
 - Only 153 out of 46 726

Top 500 Apps

Differentiating Smartphone Users by App Usage

- How many users are anonymous?
 - Only 153 out of 46 726
 - This leaves 99.67% with a unique app signature

Top 500 Apps

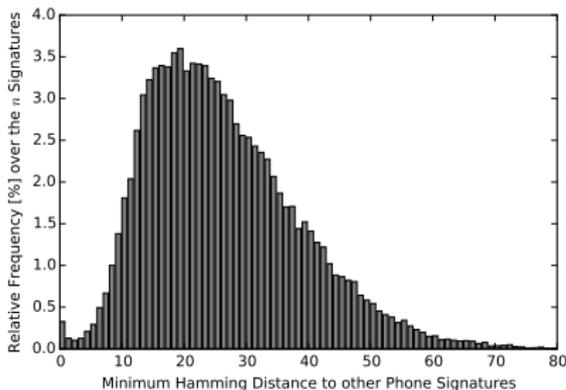
Differentiating Smartphone Users by App Usage

- How many users are anonymous?
 - Only 153 out of 46 726
 - This leaves 99.67% with a unique app signature
- How unique are these signatures?

Top 500 Apps

Differentiating Smartphone Users by App Usage

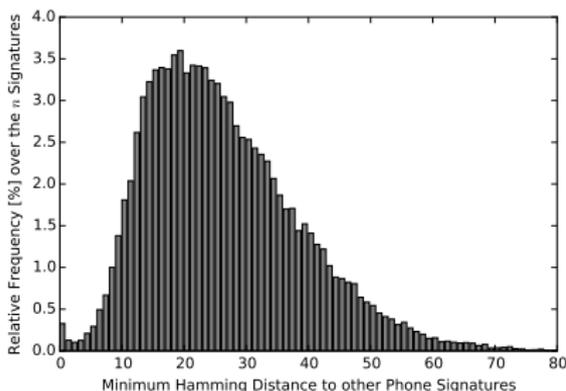
- How many users are anonymous?
 - Only 153 out of 46 726
 - This leaves 99.67% with a unique app signature
- How unique are these signatures?



Top 500 Apps

Differentiating Smartphone Users by App Usage

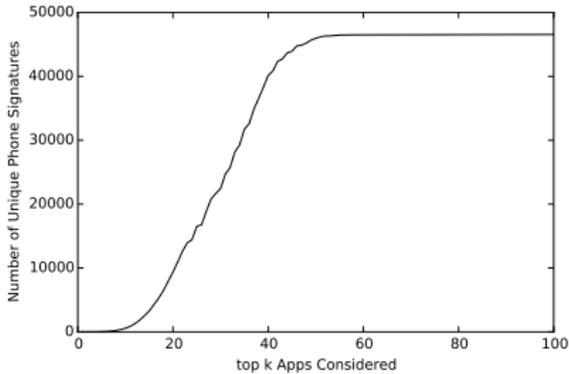
- How many users are anonymous?
 - Only 153 out of 46 726
 - This leaves 99.67% with a unique app signature
- How unique are these signatures?



- The *average minimum Hamming distance* between signatures is 25.93

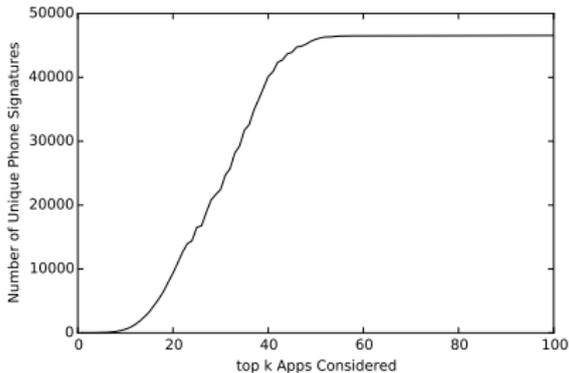
How Low Must We Go?

Differentiating Smartphone Users by App Usage



How Low Must We Go?

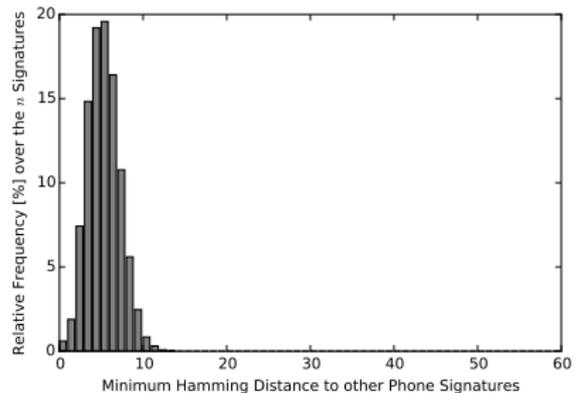
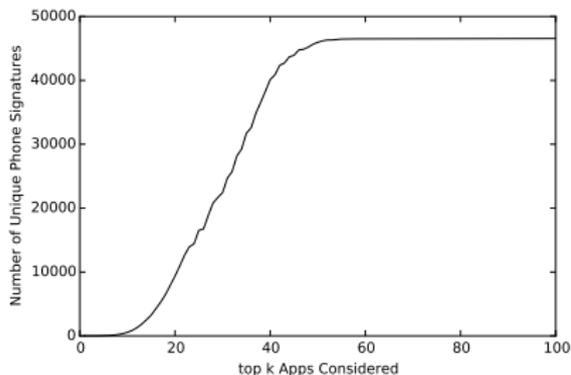
Differentiating Smartphone Users by App Usage



- \Rightarrow Almost each user has a unique signature *among the 60 most frequently used apps!*

How Low Must We Go?

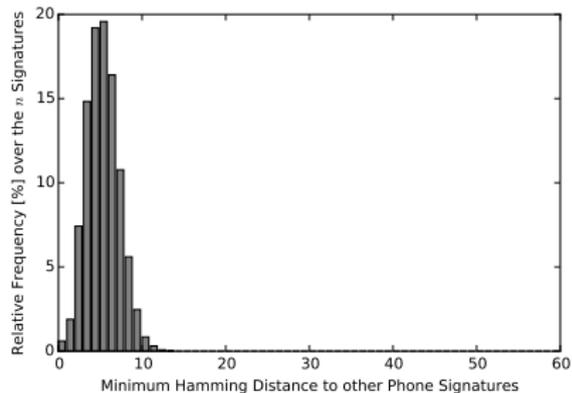
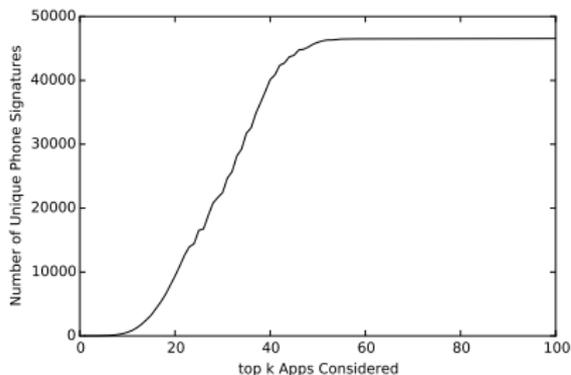
Differentiating Smartphone Users by App Usage



- \Rightarrow Almost each user has a unique signature *among the 60 most frequently used apps!*

How Low Must We Go?

Differentiating Smartphone Users by App Usage



- \Rightarrow Almost each user has a unique signature *among the 60 most frequently used apps!*

- The average minimum Hamming distance drops to 4.9

The Implications

Differentiating Smartphone Users by App Usage

- It is almost impossible to hide in the mass
 - You need to start or discontinue the use of 26 apps on average to have the signature of your neighbor (even among the 500 most popular apps)

The Implications

Differentiating Smartphone Users by App Usage

- It is almost impossible to hide in the mass
 - You need to start or discontinue the use of 26 apps on average to have the signature of your neighbor (even among the 500 most popular apps)
- There is no “average” smartphone user

The Implications

Differentiating Smartphone Users by App Usage

- It is almost impossible to hide in the mass
 - You need to start or discontinue the use of 26 apps on average to have the signature of your neighbor (even among the 500 most popular apps)
- There is no “average” smartphone user
- You are probably uniquely identifiable (*needs more work to be certain*)

The Implications

Differentiating Smartphone Users by App Usage

- It is almost impossible to hide in the mass
 - You need to start or discontinue the use of 26 apps on average to have the signature of your neighbor (even among the 500 most popular apps)
- There is no “average” smartphone user
- You are probably uniquely identifiable (*needs more work to be certain*)
 - by apps on your phone
 - by listening to your network traffic

The Implications

Differentiating Smartphone Users by App Usage

- It is almost impossible to hide in the mass
 - You need to start or discontinue the use of 26 apps on average to have the signature of your neighbor (even among the 500 most popular apps)
- There is no “average” smartphone user
- You are probably uniquely identifiable (*needs more work to be certain*)
 - by apps on your phone
 - by listening to your network traffic
 - ..even if you change your number or spoof your MAC or IMEI
 - as long as you do not change your *behavior* dramatically

Future Work

Differentiating Smartphone Users by App Usage

- Do our results generalize to *installed* apps?
 - List of installed apps is easily obtainable for any app

Future Work

Differentiating Smartphone Users by App Usage

- Do our results generalize to *installed* apps?
 - List of installed apps is easily obtainable for any app
- Can we identify users when we see them again?
 - The set of used apps might change over time

Future Work

Differentiating Smartphone Users by App Usage

- Do our results generalize to *installed* apps?
 - List of installed apps is easily obtainable for any app
- Can we identify users when we see them again?
 - The set of used apps might change over time
- How long do we need to observe the behavior for this to work?
 - a few hours?
 - a day?
 - a week?

Future Work

Differentiating Smartphone Users by App Usage

- Do our results generalize to *installed* apps?
 - List of installed apps is easily obtainable for any app
- Can we identify users when we see them again?
 - The set of used apps might change over time
- How long do we need to observe the behavior for this to work?
 - a few hours?
 - a day?
 - a week?
- Does all of this hold for iPhones, as well?