CS 528 Mobile and Ubiquitous Computing

Lecture 9b: Mobile Security and Mobile Measurements

Emmanuel Agu

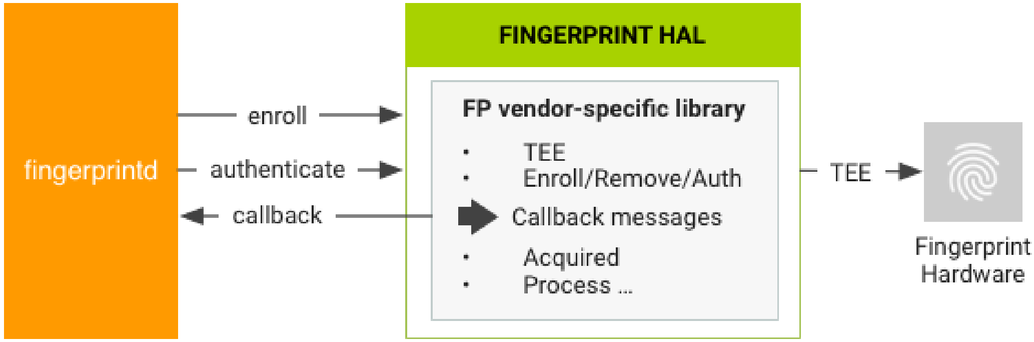
Authentication using Biometrics

Biometrics

- Passwords tough to remember, manage
- Many users have simple passwords (e.g. 1234) or do not change passwords
- Biometrics are unique physiological attributes of each person
  - Fingerprint, voice, face
- Can be used to replace passwords
  - No need to remember anything. Just be you. Cool!!

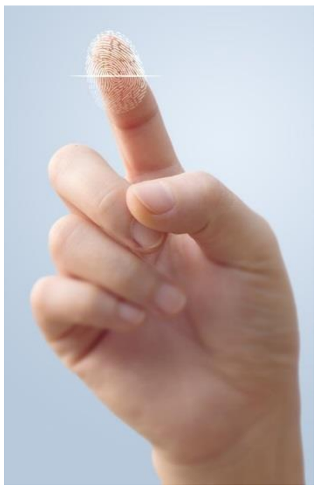
Android Biometric Authentication: Fingerprints

- **Fingerprint:** On devices with fingerprint sensor, users can enroll multiple fingerprints for unlocking device



Samsung Pass: More Biometrics

- **Samsung pass:** Fingerprint + Iris scan + facial recognition



- Probably ok to use for facebook, social media
- Spanish bank BBVA’s mobile app uses biometrics to allow login without username + password
- Bank of America: pilot testing iris authentication since Aug 2017

Continuous Passive Authentication using Behavioral Biometrics

User Behavior as a Biometric

- User behaviors patterns are unique personal features. E.g
  - Each person's daily location pattern (home, work, places) + times
  - Walk pattern
  - Phone tilt pattern

- **General idea:** Continuously authenticate user as long as they behave like themselves

- If we can measure user behavior reliably, this could enable **passive authentication**

BehavioMetrics

Ref: Zhu *et al,* Mobile Behaviometrics: Models and Applications

- Derived from Behavioral Biometrics
  - Behavioral: the way a human subject behaves
  - Biometrics: technologies and methods that measure and analyzes biological characteristics of the human body
    - Fingerprints, eye retina, voice patterns
- BehavioMetrics:
  - Measurable behavior to recognize or verify a human's identity

Mobile Sensing → BehavioMetrics

- Accelerometer
  - Activity & movement pattern, hand trembling, driving style
  - sleeping pattern
  - Activity level, steps per day, calories burned
- Motion sensors, WiFi, Bluetooth
  - Indoor position and trajectory.
- GPS
  - outdoor location, geo-trace, commuting pattern
- Microphone, camera
  - From background noise: activity, type of location.
  - From voice: stress level, emotion
  - Video/audio: additional contexts
- Keyboard, taps, swipes
  - User interactions, tasks .....

- *Network Factors*
- *Personal Factors*
- *Behavioral Factors*
- *Application Factors*

BehavioMetrics → Security

- Track smartphone user behavior using sensors
- Continuously extract and classify features from sensors = Detect contexts, personal behavior features (pattern classification)
- Generate unique pattern for each user
- **Trust score:** How similar is today's behavior to user's typical behavior
- Trigger authentication schemes with different levels of authentication based on trust score

[31,271,37] [37,281,42] [37,276,47] [42,271,47] [42,266,53] [58,271,47] [53,271,47] [74,271,42] . . .

⇓

CZ DG GI FK C BI CS DC HQ BX FI FI BX FI O . . .

Continuous n-gram Model

- User activity at time *i* depends only on the last *n-1* activities
- Sequence of activities can be predicted by *n* consecutive activities in the past

$$P(l_i | l_{i-n+1}, l_{i-n+2}, \ldots, l_{i-1}) \quad \text{or} \quad P(l_i | l_{i-n+1}^{i-1})$$

- Maximum Likelihood Estimation from training data by counting:

$$P_{\text{MLE}}(l_i | l_{i-n+1}^{i-1}) = \frac{C(l_{i-n+1}, \ldots, l_{i-1}, l_i)}{C(l_{i-n+1}, \ldots, l_{i-1})}$$

- MLE assign zero probability to unseen n-grams

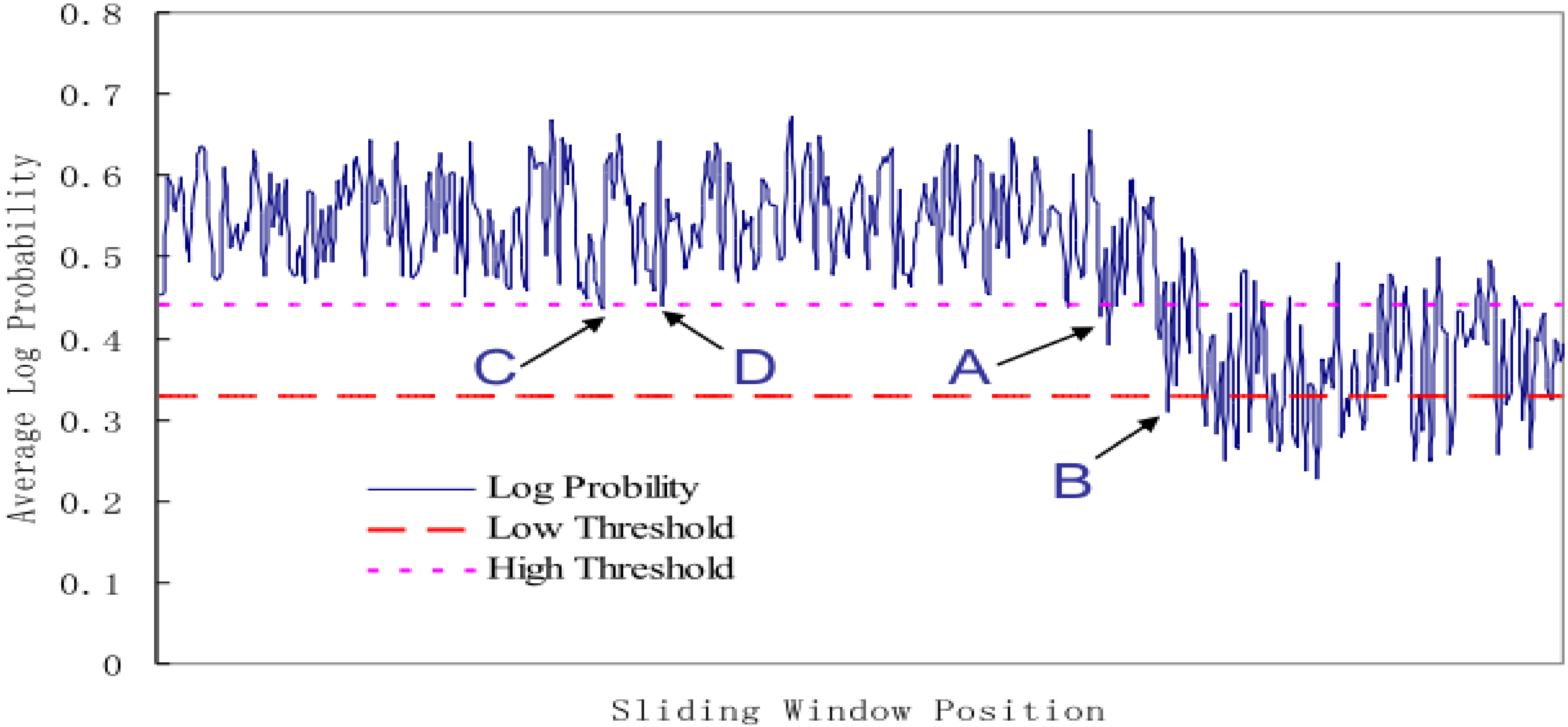
Classification

- Build *M* BehavioMetrics models $P_0, P_1, P_2, \ldots, P_{M-1}$
  - Genders, age groups, occupations
  - Behaviors, activities, actions
  - Health and mental status

- Classification problem formulated as

$$\hat{u} = \underset{m}{\operatorname{argmax}}\, P(L, m) = \underset{m}{\operatorname{argmax}} \sum_{i=1}^{N} \log P_m(l_i | l_{i-n+1}^{i-1})$$

Anomaly Detection Threshold



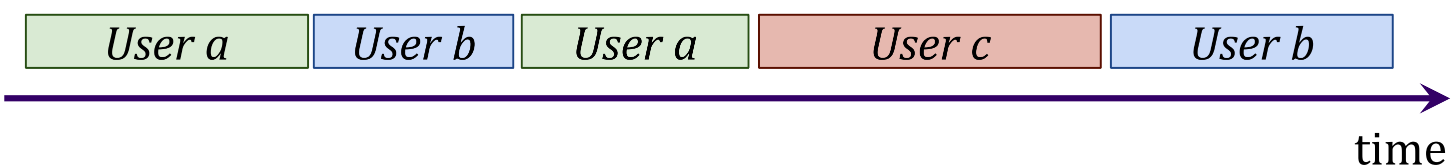
Behavioral Biometrics Issues:
Shared Devices

BehavioMetric Issues: Multi-Person Use

- Many mobile devices are shared by multiple people
  - Classifier trained using person A's data cannot detect Person B

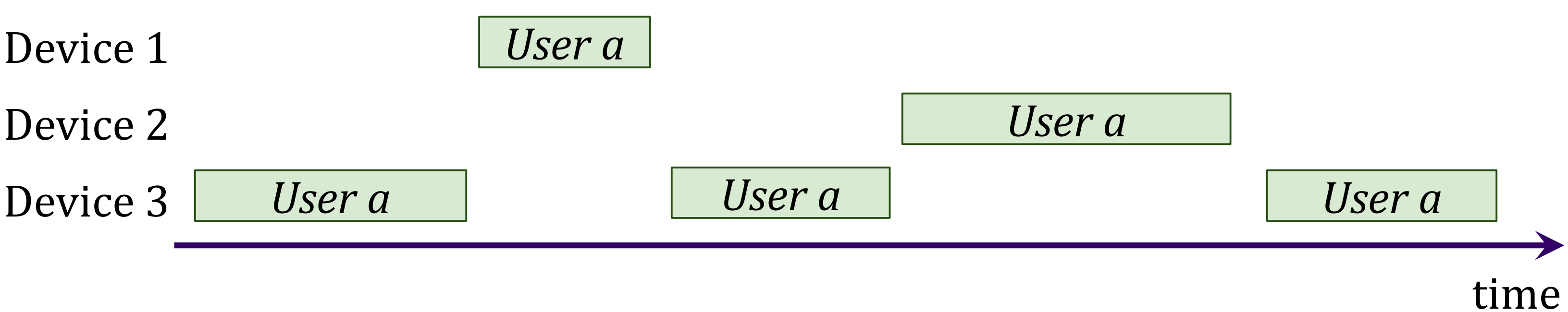
- **Question:** How to distinguish when person A vs person B using the shared device

- How to segment the activities on a single device to those of multiple users?



BehavioMetric Issues: Multi-Device Use

- Many people have multiple mobile devices
  - Classifier trained on device 1 (e.g. smartphone) may not detect behavior on device 2 (e.g. smartwatch)
- **Question:** How to match same user's session on multiple devices
    - **E.g.** Use Classifier trained on smartphone to recognize user on smartwatch
- How to match user's activity segments on different devices?

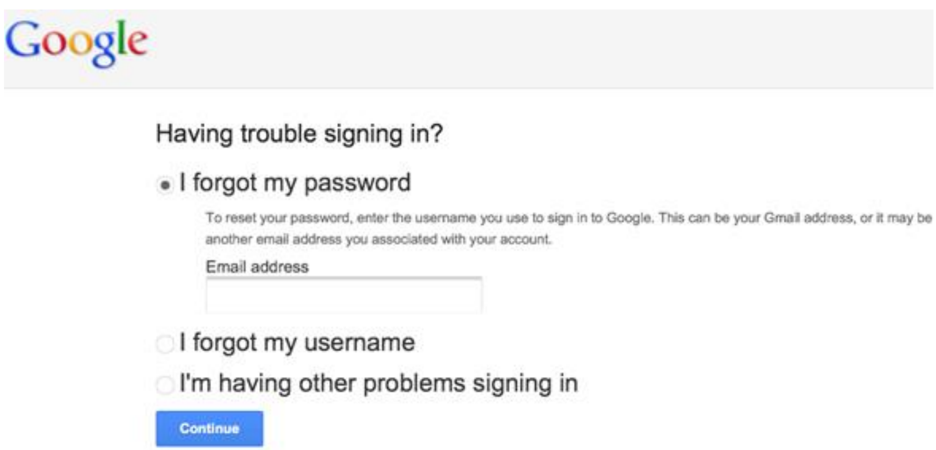


ActivPass

ActivPass

S. Dandapat, S Pradhan, B Mitra, R Choudhury and N Ganguly, ActivPass: Your Daily Activity is Your Password, in Proc CHI 2015

- Passwords are mostly secure, simple to use but have issues:
  - Simple passwords (e.g. 1234): easy to crack
  - Secure passwords hard to remember (e.g. $emime)$@(*$@)9)
  - Remembering passwords for different websites even more challenging
  - Many people use same password on different websites (dangerous!!)



ActivPass

S. Dandapat, S Pradhan, B Mitra, R Choudhury and N Ganguly, ActivPass: Your Daily Activity is Your Password, in Proc CHI 2015

- Unique human biometrics being explored
- **Explicit biometrics:** user actively makes input
  - E.g. finger print, face print, retina scan, etc
- **Implicit biometrics:** works passively, user does nothing explicit to be authenticated.
  - E.g. unique way of walk, typing, swiping on screen, locations visited daily
- **This paper:** smartphone soft sensors as biometrics: calls, SMS, contacts, etc
- **Advantage of biometrics:** simple, no need to remember anything

ActivPass Vision

- **Observation:** rare events are easy to remember, hard to guess
  - E.g. A website user visited this morning that they rarely visits
    - User went to CNN.com today for the first time in 2 years!
  - Got call from friend I haven't spoken to in 5 years for first time today

- **Idea:** Authenticate user by quizzing them to confirm rare (outlier) activities
  - What is caller's name from first call you received today?
  - Which news site did you not visit today? (CNN, CBS, BBC, Slashdot)?

ActivPass Vision

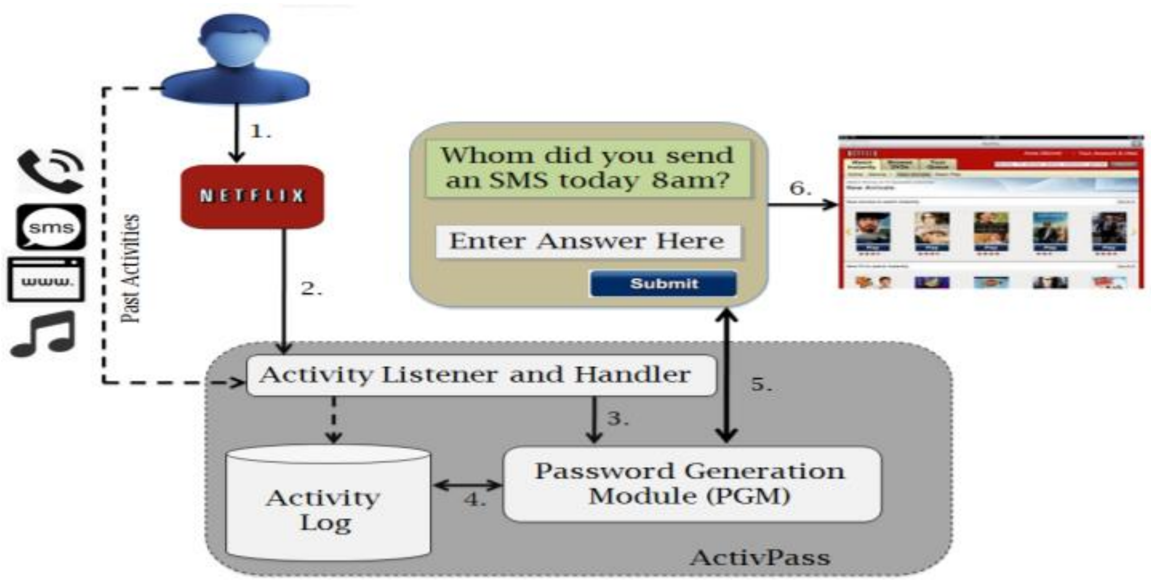
- Authentication questions based on outlier (rare) activities generated from:
  - Call logs
  - SMS logs
  - Facebook activities
  - Browser history

ActivPass Envisioned Usage Scenarios

- Replace password hints with Activity questions when password lost
- Combine with regular password (soft authentication mechanism)
- Prevent password sharing.
  - E.g. Bob pays for Netflix, shares his login details with Alice

How ActivPass Works

- Activity Listener runs in background, logs
  - Calls, SMS, web pages visited, etc
- When user launches an app:
  - Password Generation Module (PGM) creates *n* password questions based on logged data
  - If user can answer *k* of password questions correctly, app is launched!



ActivPass Vision

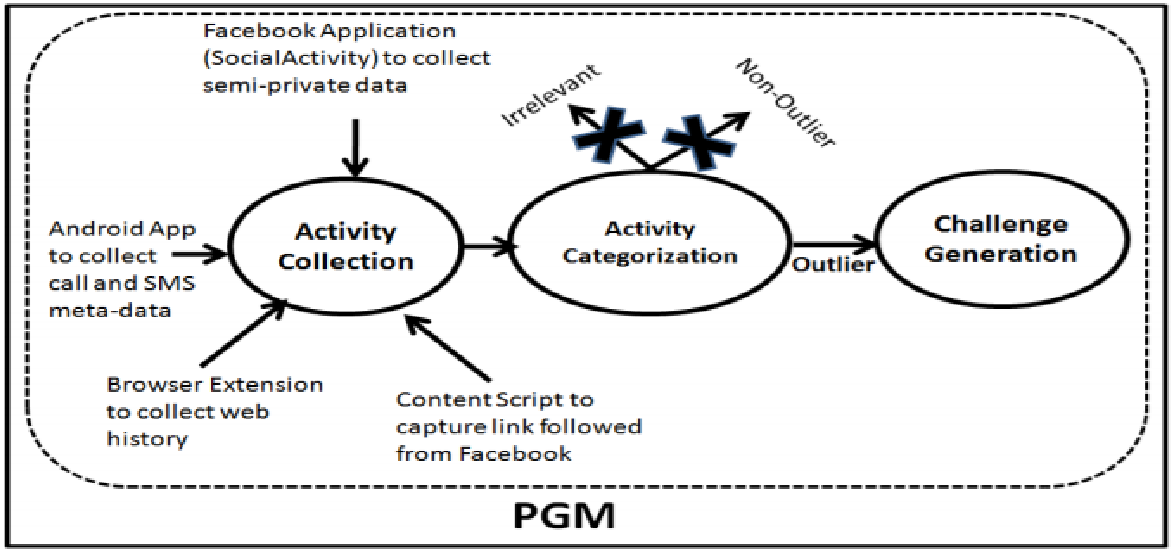
- User can customize
  - Number of questions asked,
  - What fraction of questions *k* must be answered correctly
  - Question format
  - Activity permissions

Question formats	Example questions asked
Binary	Have you received a call from Alice at around 10 pm on 19/09/2014?
MCQ	Please write the options of the links you visited,this week in comma separated way (Ex: A, B): A. CNN; B. BBC; C. SKY News; D. Reuters
Text	Whom did you call at around 7 pm on 17/09/2014 ? Hint: (Al*)

- Paper investigates ActivPass utility by conducting user studies

How ActivPass Works

- Periodically retrieves logs in order to classify them using **Activity Categorization Module**
  - Tries to find outliers in the data. E.g. Frequently visited pages vs rarely visited web pages



ActivPass: Types of Questions Asked Vs Data Logged

	Range of questions asked
Facebook	1) Profiles visited by the user. 2) Groups the user is a member of. 3) A person with whom user had a chat.
Web	1) Titles of the web-pages visited by the user.
Call	1) A person whom the user called. 2) A person who called the user.
SMS	1) A person whom the user sent an SMS. 2) A person who sent an SMS to the user.
Audio	1) The tune/tone used by the user as an alarm. 2) The tune/tone used by the user as her ring-tone. 3) The audio files downloaded by the user.

Source	**Details of data collected**
SMS	Time, Receiver/Sender Name
Call	Time, Type (incoming, outgoing), Name of other person, Duration
Audio	Title of Music added in this week, Alarm tone, Ring tone
Web	URL, Time of visit
Link visited from Facebook	URL, Time of visit
Facebook Group	Name of Private (secret and closed) groups
Facebook Pages	Name of pages created by user
Facebook Profile	Name of Facebook friends of user
Facebook Message	Time (in milliseconds from epoch), Name of other person, Msg Id, Thread Id

ActivPass: Evaluation

- Over 50 volunteers given 20 questions:
  - Avg. recall rate: 86.3% ± 9.5 (user)
  - Avg guessability: 14.6% ± 5.7 (attacker)
- Devised Bayesian estimate of challenge given *n* questions where *k* are required
- Tested on 15 volunteers
  - Authenticates correct user 95%
  - Authenticates imposter 5.5% of the time (guessability)

n	k	Authentic user	Impostor
4	4	0.554	0.0004
4	3	0.906	0.011
4	2	0.989	0.1043
4	1	0.998	0.468
3	3	0.642	0.0031
3	2	0.948	0.0577
3	1	0.996	0.3771
2	2	0.745	0.0213
2	1	0.981	0.2707

Optimal *n, k*

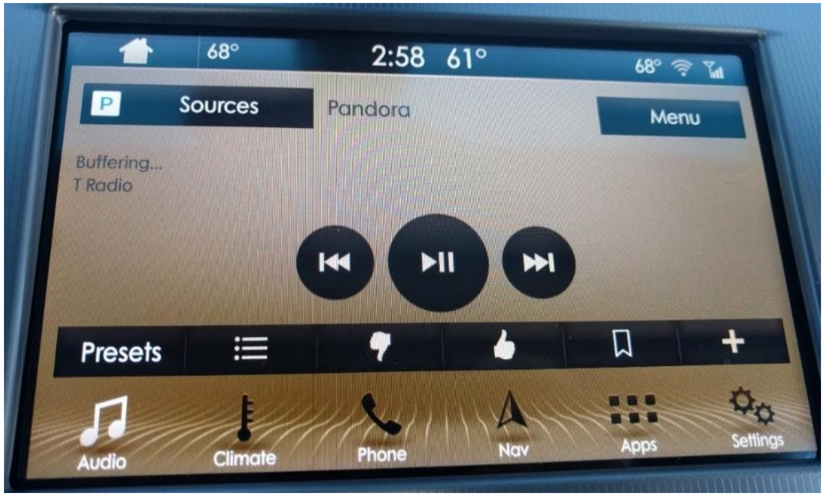
Maximize

Minimize

Smartphones + IoT Security Risks

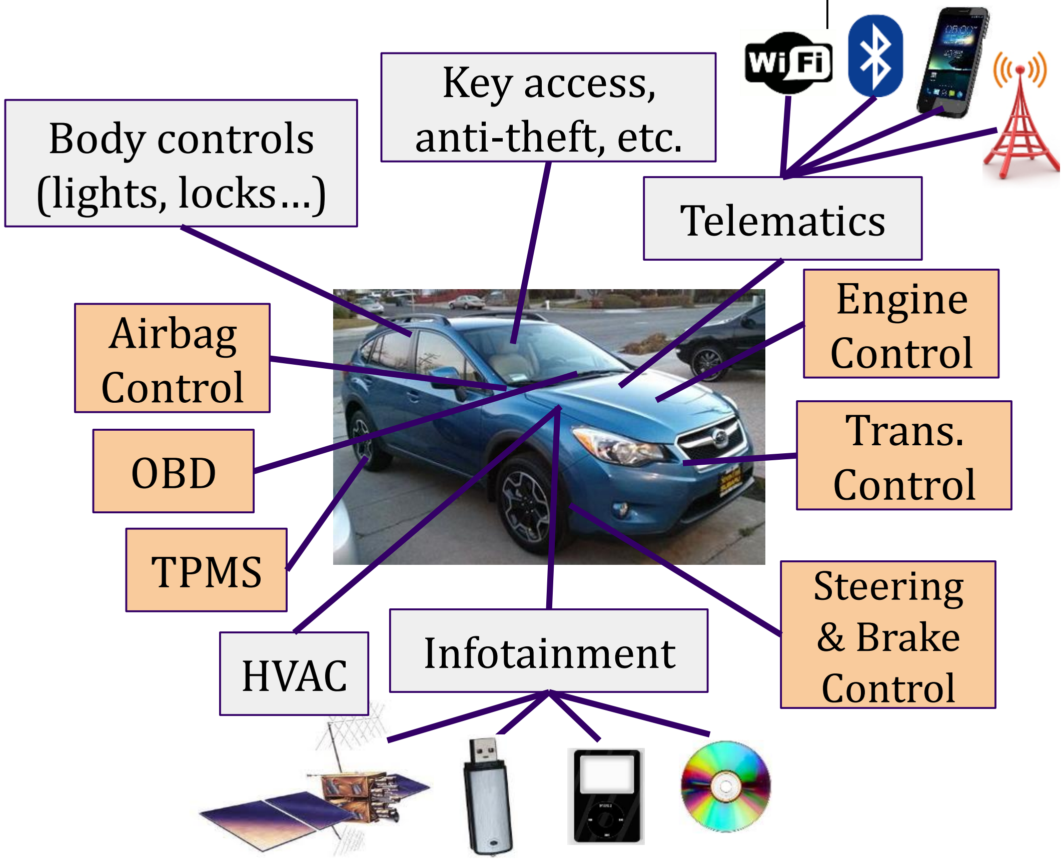
Cars + Smartphones → ?

- Many new vehicles come equipped with smartphone integration / capabilities in the infotainment system (Android Auto!)



Smartphones that Drive

- If a mobile app gets access to a vehicle's infotainment system, is it possible to get access to (or even to control) driving functionality?



Smart Vehicle Risks

- Many of the risks and considerations that we discussed in this course can be applied to smart vehicles and smartphone interactions
- However, many more risks come into play because of the other functionality that a car has compared to a smartphone

CS 528 Mobile and Ubiquitous Computing
Secure Mobile Software Development (SMSD)

Emmanuel Agu

Secure Mobile Software Development Modules

Introduction

- Many Android smartphones compromised because users download malicious software disguised as legitimate apps
- Malware vulnerabilities can lead to:
  - Stolen credit card numbers, financial loss
  - Stealing user's contacts, confidential information
- Frequently, unsafe programming practices by software developers expose vulnerabilities and back doors that hackers/malware can exploit
- Examples:
  - Attacker can send invalid input to your app, causing confidential information leakage

Secure Mobile Software Development (SMSD)

- **Goal:** Teach mobile (Android) developers about backdoors, reduce vulnerabilities in shipped code

- SMSD:
  - Hands-on, engaging labs to teach concepts, principles
  - Android plug-in: Highlights, alerts Android coder about vulnerabilities in their code
  - Quite useful

SMSD: 8 Modules

- M0: Getting started
- M1: Data sanitization for input validation
- M2: Data sanitization for output encoding
- M3: SQL injections
- M4: Data protection
- **M5: Secure inter-process communication (IPC)**
- M6: Secure mobile databases
- M7: Unintended data leakage
- **M8: Access control**

- You should
  - Pre-Survey
  - **Lab:** Go through M5, M8
  - Post-survey afterwards

M5 & M8 Overview

- M5: Intra-app IPC vulnerabilities
- 2 security loopholes
  - **Intent Eavesdropping:** Malicious app can **receive** intent not meant for it
  - **Intent Spoofing:** Malicious app inserts (**send**) undesired behavior into a component using the implicit intent

- M8: Inter-App Secure IPC vulnerabilities
  - Malicious app can exploit security loophole in Broadcast Receivers to intercept valuable information

Important: This Lab REPLACES Worst Quiz

- Counts as quiz 6
- I will drop your worst quiz and replace it with score from SMSD
- Basically, I will use your best 5 scores
- Just do this lab online,
- Due 11.59, Friday, December 14, 2018

Mobile Measurements: Android Users in China

Introduction

Huoran Li et al., "Characterizing Smartphone Usage Patterns from Millions of Android Users"
Internet Measurement Conference (IMC) 2015

- Understanding user behaviors while using mobile apps is critical. Why?
  - App stores can build better recommender systems
  - Developers can better understand why users like certain apps
- This paper presents results of a comprehensive measurement study to investigate smartphone user patterns
- Sample questions addressed:
  - Characterize app popularity among millions of users?
  - Understand how mobile users choose and manage apps?
  - Type and amount of network traffic generated by various apps
  - Investigate economic factors affect app selection and network behavior?

Dataset

- Gathered from Wandoujia, leading Android App Store in China

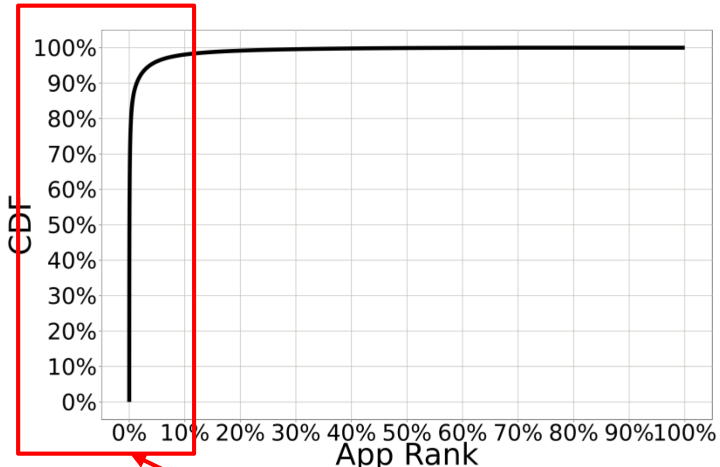
- Wandoujia:
  - Over 250 million users in 2015
  - All apps are free

- 1 month of data gathering
  - Over 8 million unique users
  - Over 260,172 unique apps in dataset

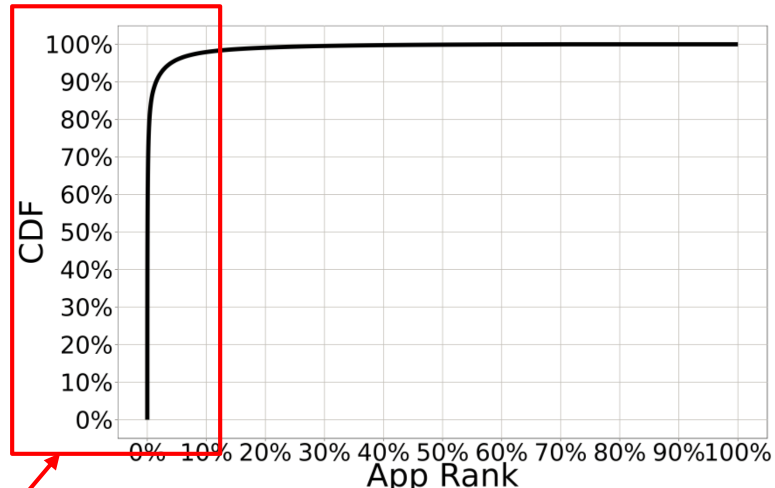
App Popularity Metrics

- No. of downloads of each app
- No. of unique devices that download each app;
- Total data traffic generated by each app;
- Total access time users spend interacting with each app.

App Popularity: Downloads & Unique Subscribers



Percentage of Downloads against App Rank



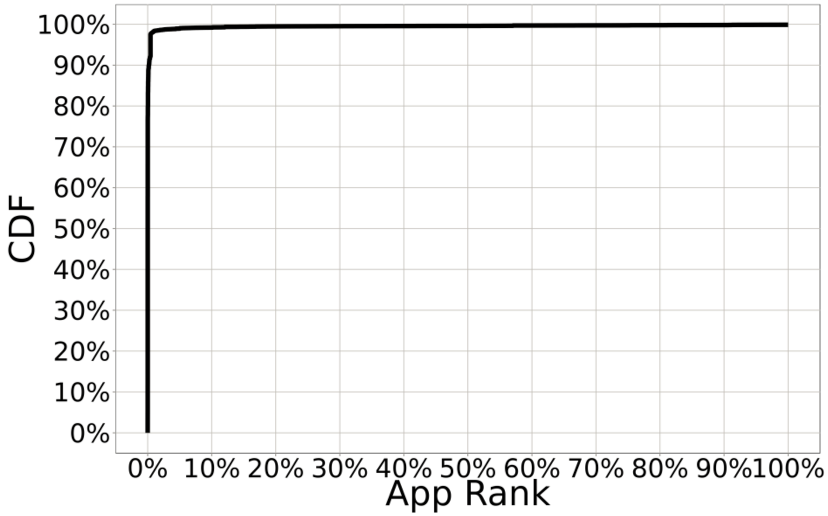
Percentage of Unique Subscribers against App Rank

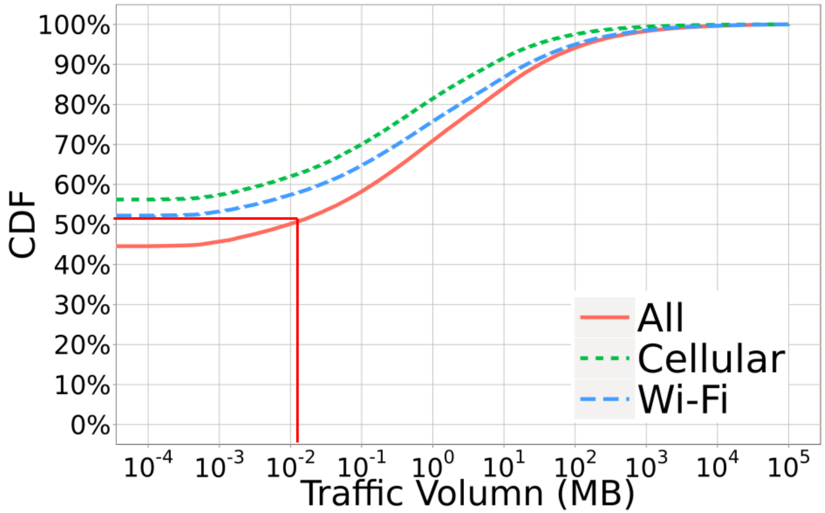
Top 10% of apps get over 99% of the downloads and Unique subscribers

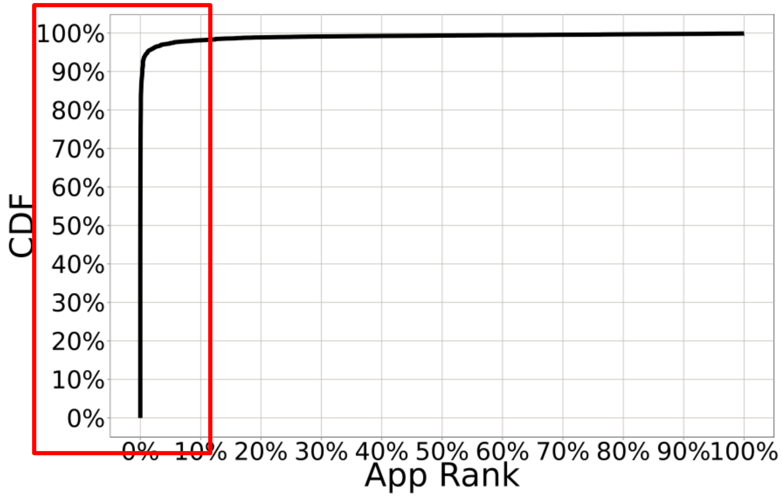
App Popularity: Network Traffic

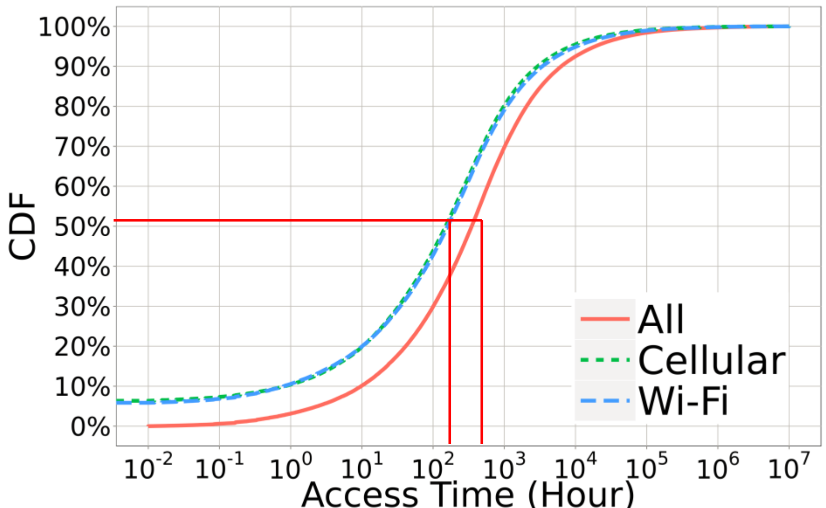
Top-ranked 10% of apps generates over 99% of network traffic

97% apps consume < 100 MB traffic per 1 month
95% of apps are used less than 100 hours/mo



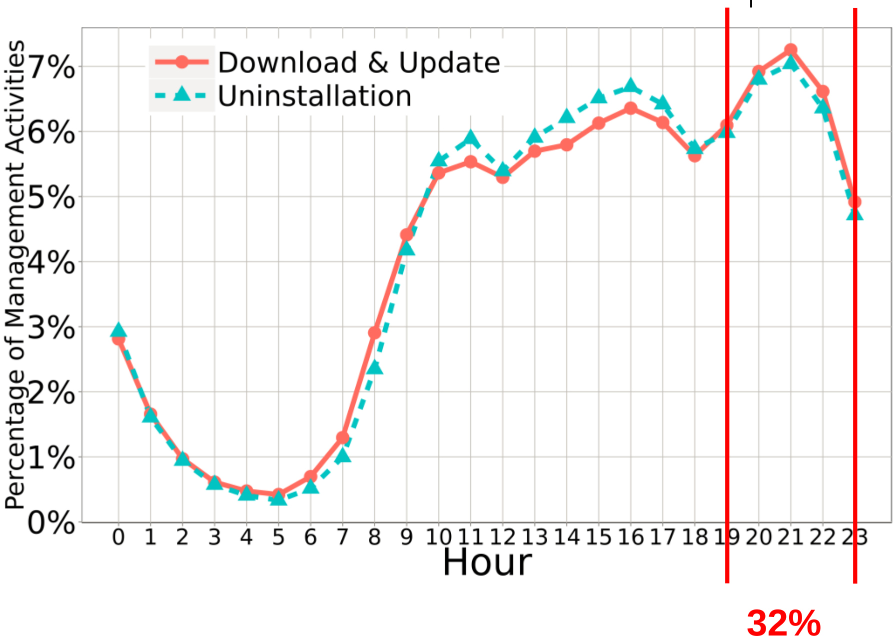






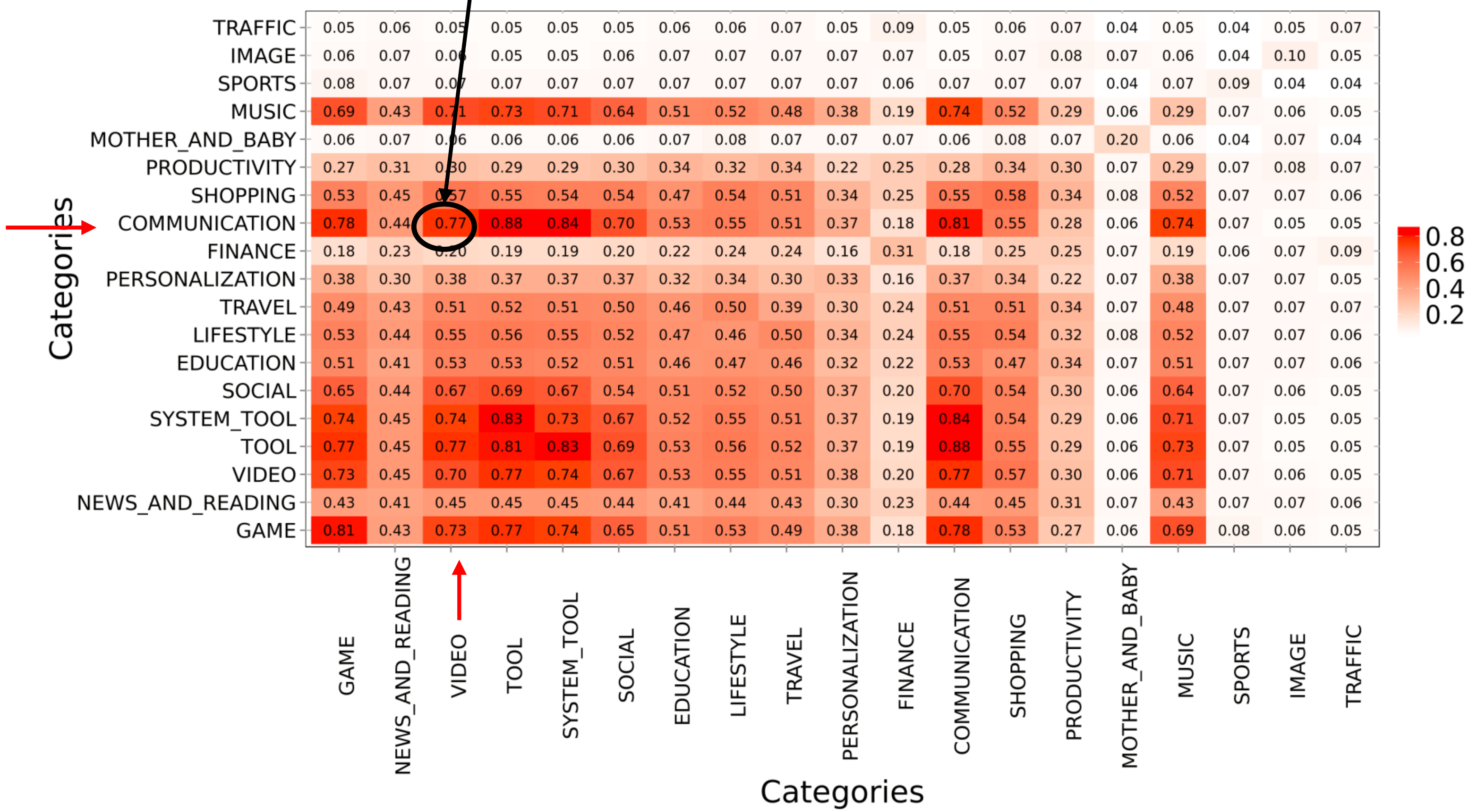
App Management & Installation Patterns

- About 32% of app downloading and updating activities performed between 7:00 pm to 11:00 pm (at night)



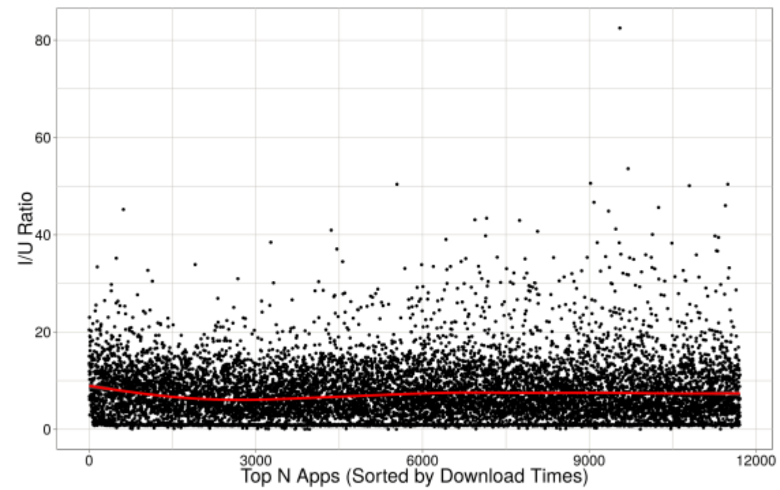
App Co-Occurrence of App Categories

- Gives sense of apps users like to use together
- E.g. Many users like to share video = high co-occurrence of video + communication apps (E.g. share videos on whatsapp)

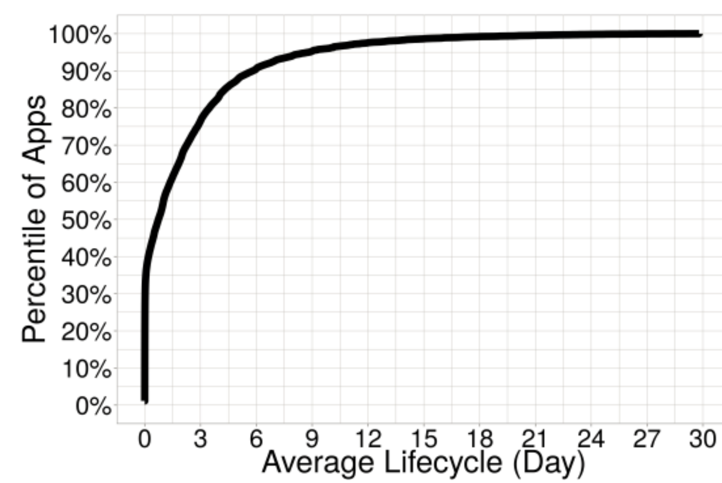


App Uninstallation Patterns

- **I/U ratio:** No. of Installations/No. Uninstallation
  - E.g. I/U = 8 => 1 out of 8 users who download the app uninstall it



- Users react quickly to disliked apps
- Of all apps that are uninstalled
  - 40% are uninstalled within 1 day
  - 93% are uninstalled within 1 week



Data Traffic Patterns

- Video apps consume over 81% of Wi-Fi traffic and 28% of cellular traffic
- Users are more likely to lauch video apps on WiFi

Table 1: Chosen Top Apps by Category.

App Category	Apps	Users (10^6 devices)	Downloads (10^6 times)	Traffic (GB)	Access-Time (10^7 hours)	*C*-Traffic	*C*-Time	*W*-Traffic	*W*-Time
GAME	1,227	3.87	15.15	13,669.71	0.38	2.98%	5.19%	0.76%	6.39%
NEWS_AND_READING	274	1.17	1.97	13,143.17	0.23	3.11%	2.91%	0.72%	3.95%
VIDEO	238	2.86	6.52	1,196,978.79	0.38	28.41%	1.42%	81.08%	10.54%
TOOL	227	3.84	9.43	77,329.87	0.68	15.63%	10.79%	4.40%	9.46%
SYSTEM_TOOL	217	3.37	7.54	34,012.16	0.25	3.05%	3.37%	2.17%	4.24%
SOCIAL	188	2.18	4.01	35,926.76	0.35	8.96%	4.77%	1.94%	5.66%
EDUCATION	172	1.68	2.98	13,893.55	0.34	1.46%	5.35%	0.87%	4.71%
LIFESTYLE	156	1.68	2.85	2,388.59	0.07	0.72%	1.00%	0.12%	1.06%
TRAVEL	111	1.62	2.75	8,182.24	0.03	0.78%	0.53%	0.52%	0.25%
PERSONALIZATION	104	1.49	3.68	7,426.38	0.86	0.85%	12.03%	0.46%	13.67%
FINANCE	99	0.32	0.50	382.60	0.02	0.13%	0.24%	0.02%	0.26%
COMMUNICATION	85	4.09	8.45	54,394.71	2.85	24.74%	49.01%	2.26%	35.26%
SHOPPING	78	1.57	3.00	21,808.51	0.07	3.16%	0.65%	1.32%	1.60%
PRODUCTIVITY	75	0.76	1.17	2,712.50	0.01	0.18%	0.17%	0.18%	0.26%
MOTHER_AND_BABY	48	0.10	0.15	525.72	0.01	0.07%	0.04%	0.03%	0.12%
MUSIC	43	2.33	3.39	49,540.12	0.17	5.66%	2.47%	3.08%	2.49%
SPORTS	27	0.31	0.36	61.40	0.00	0.02%	0.05%	0.00%	0.04%
IMAGE	23	0.14	0.17	801.64	0.00	0.06%	0.01%	0.05%	0.03%
TRAFFIC	14	0.10	0.12	78.10	0.00	0.02%	0.03%	0.00%	0.01%

The users, downloads, traffic, and access time are all computed by aggregating the data of each app in the category
The percentile of *W*-Traffic (*C*-Traffic) and *W*-Time (*C*-Time) refer to the data traffic and foreground access time over Wi-Fi (W) and cellular (C) network, respectively.

Data Traffic of Foreground and Background

- App categories with high traffic:
  - VIDEO: prefetching of videos
  - SYSTEM_TOOL: Anti-virus updating
  - GAMES: Embedded ads
- < 2% of network access time in foreground, 98% in background
  - Many apps keep long-lived background TCP/IP connections. Secret downloads. Hmm…

Table 2: Network Summary by App Category

App Category	C-Traffic (B)	W-Traffic (B)	C-Traffic (F)	W-Traffic (F)	C-Time (B)	W-Time (B)	C-Time (F)	W-Time (F)
VIDEO	0.81%	45.13%	1.28%	52.78%	42.62%	56.66%	0.10%	0.63%
TOOL	8.16%	39.13%	9.56%	43.14%	48.57%	50.42%	0.57%	0.43%
COMMUNICATION	12.42%	15.90%	27.48%	44.20%	48.01%	46.85%	3.15%	1.99%
MUSIC	4.35%	35.19%	5.67%	54.80%	49.23%	50.09%	0.36%	0.32%
SOCIAL	7.26%	20.65%	14.63%	57.47%	48.43%	50.41%	0.57%	0.59%
SYSTEM_TOOL	5.07%	51.57%	2.80%	40.55%	50.02%	49.48%	0.23%	0.26%
SHOPPING	3.29%	17.09%	9.42%	70.21%	43.34%	56.42%	0.08%	0.17%
EDUCATION	3.76%	39.38%	5.46%	51.40%	45.57%	52.83%	0.90%	0.69%
GAME	10.34%	43.11%	8.80%	37.74%	48.13%	51.34%	0.26%	0.28%
NEWS_AND_READING	5.91%	24.64%	14.83%	54.62%	43.43%	55.25%	0.60%	0.71%

W and C refer to Wi-Fi and Cellular, respectively.
B refers to background and F refers to foreground.

Device Model Clustering

- Device model are Moto G5, Samsung galaxy 6, etc
- 96% device models have less than 500 users

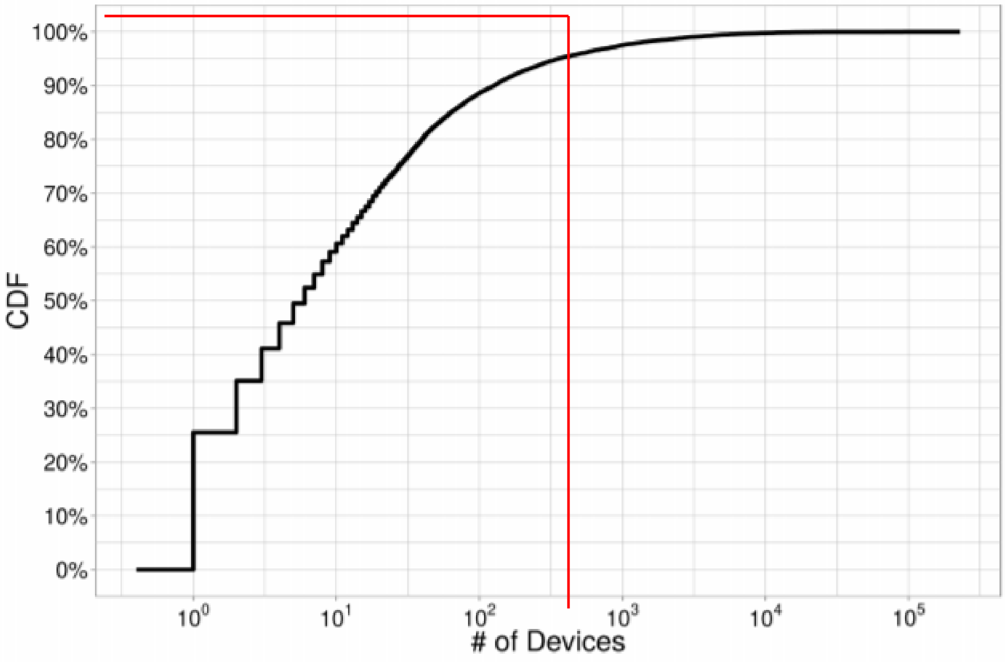
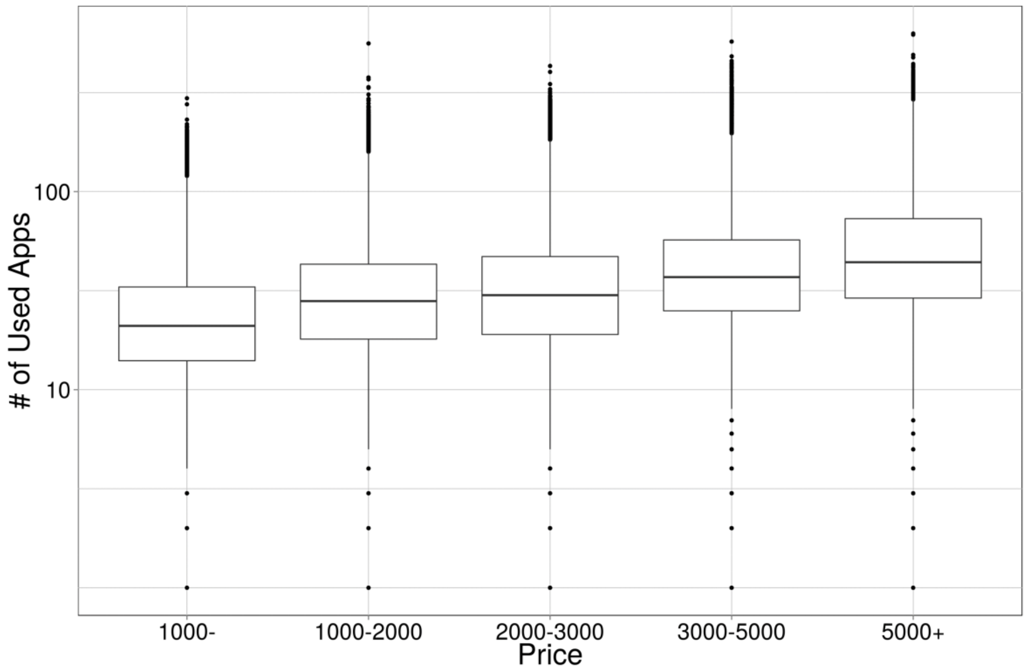


Figure 10: CDF for Number of Users of Device Models

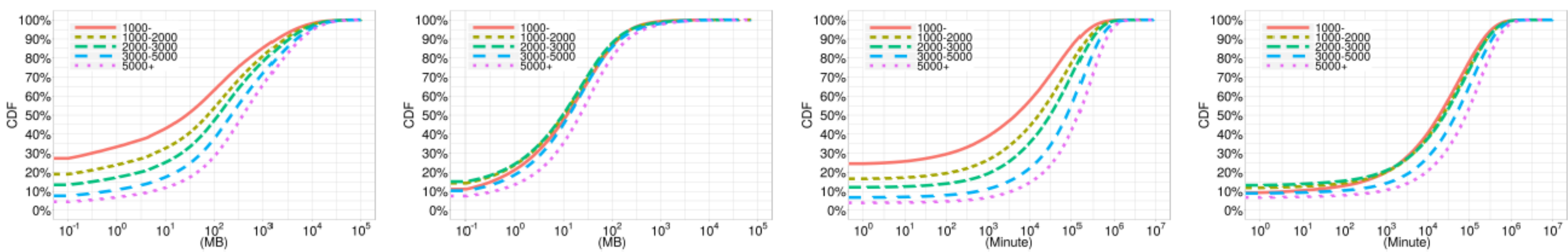
Apps Installed on Various Device Groups

- Higher priced devices have more apps installed, maybe because
  a) More RAM, better CPU, hardware, etc
  b) Bigger manufacturers who pre-install apps (bloatware)



Network Activity & App Preference Among Device Groups

- Wi-Fi usage correlated with device model prices
  - i.e. higher priced devices consume more Wi-Fi traffic



(a) Wi-Fi traffic volume (b) Cellular traffic volume (c) WiFi time (d) Cellular time

Figure 12: Network Activity Distribution among User Groups

- Also, different groups of devices (based on price) had different app preferences (e.g. browser, eBook, etc)

Study Limitations

Limitations:

- Dataset was from 1 app marketplace in China
- Users are mostly Chinese.
- Other regions may be different
- Need to look at other groups to get complete picture
- Study and analysis was on 1 month of usage data