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: Fingerprint

- **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
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
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

$$\hat{u} = \arg\max_m P(L, m) = \arg\max_m \sum_{i=1}^{N} \log P_m(l_i | l_{i-1}^{i-1})$$
Anomaly Detection Threshold

![Graph showing Anomaly Detection Threshold](image)

- **C**: Log Probability
- **D**: Low Threshold
- **A**: High Threshold
- **B**: Sliding Window Position

**Average Log Probability**

- 0.8
- 0.7
- 0.6
- 0.5
- 0.4
- 0.3
- 0.2
- 0.1
- 0.0
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
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!!)
Unique human biometrics being explored

Explicit biometrics: user actively makes input
- E.g. fingerprint, 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

<table>
<thead>
<tr>
<th>Question formats</th>
<th>Example questions asked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary</td>
<td>Have you received a call from Alice at around 10 pm on 19/09/2014?</td>
</tr>
<tr>
<td>MCQ</td>
<td>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</td>
</tr>
<tr>
<td>Text</td>
<td>Whom did you call at around 7 pm on 17/09/2014? Hint: (Al*)</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Range of questions asked</th>
<th>Details of data collected</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Facebook</strong></td>
<td></td>
</tr>
<tr>
<td>1) Profiles visited by the user.</td>
<td>Time, Receiver/Sender Name</td>
</tr>
<tr>
<td>2) Groups the user is a member of.</td>
<td></td>
</tr>
<tr>
<td>3) A person with whom user had a chat.</td>
<td></td>
</tr>
<tr>
<td><strong>Web</strong></td>
<td></td>
</tr>
<tr>
<td>1) Titles of the web-pages visited by the user.</td>
<td>Time, Type (incoming, outgoing), Name of other person, Duration</td>
</tr>
<tr>
<td><strong>Call</strong></td>
<td></td>
</tr>
<tr>
<td>1) A person whom the user called.</td>
<td></td>
</tr>
<tr>
<td>2) A person who called the user.</td>
<td></td>
</tr>
<tr>
<td><strong>SMS</strong></td>
<td></td>
</tr>
<tr>
<td>1) A person whom the user sent an SMS.</td>
<td>Title of Music added in this week, Alarm tone, Ring tone</td>
</tr>
<tr>
<td>2) A person who sent an SMS to the user.</td>
<td>URL, Time of visit</td>
</tr>
<tr>
<td><strong>Audio</strong></td>
<td></td>
</tr>
<tr>
<td>1) The tune/tone used by the user as an alarm.</td>
<td>URL, Time of visit</td>
</tr>
<tr>
<td>2) The tune/tone used by the user as her ring-tone.</td>
<td></td>
</tr>
<tr>
<td>3) The audio files downloaded by the user.</td>
<td></td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( k )</th>
<th>Authentic user</th>
<th>Impostor</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>0.554</td>
<td>0.0004</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.906</td>
<td>0.011</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.989</td>
<td>0.1043</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.998</td>
<td>0.468</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.642</td>
<td>0.0031</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.948</td>
<td>0.0577</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.996</td>
<td>0.3771</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.745</td>
<td>0.0213</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.981</td>
<td>0.2707</td>
</tr>
</tbody>
</table>

Optimal \( n, k \)
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.
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
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

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)

<table>
<thead>
<tr>
<th>Categories</th>
<th>TRAFFIC</th>
<th>IMAGE</th>
<th>SPORTS</th>
<th>MUSIC</th>
<th>MOTHER_AND_BABY</th>
<th>PRODUCTIVITY</th>
<th>SHOPPING</th>
<th>COMMUNICATION</th>
<th>FINANCE</th>
<th>PERSONALIZATION</th>
<th>TRAVEL</th>
<th>LIFESTYLE</th>
<th>EDUCATION</th>
<th>SOCIAL</th>
<th>SYSTEM_TOOL</th>
<th>TOOL</th>
<th>VIDEO</th>
<th>NEWS_AND_READING</th>
<th>GAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAFFIC</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>IMAGE</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>SPORTS</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>MUSIC</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>MOTHER_AND_BABY</td>
<td>0.6</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PRODUCTIVITY</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>SHOPPING</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>COMMUNICATION</td>
<td>0.8</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>FINANCE</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>PERSONALIZATION</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>TRAVEL</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>LIFESTYLE</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>EDUCATION</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>SOCIAL</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
<td>0.6</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>SYSTEM_TOOL</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>TOOL</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>VIDEO</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.7</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td>NEWS_AND_READING</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>GAME</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>
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 launch video apps on WiFi.

### Table 1: Chosen Top Apps by Category.

<table>
<thead>
<tr>
<th>App Category</th>
<th>Apps</th>
<th>Users (10^6 devices)</th>
<th>Downloads (10^6 times)</th>
<th>Traffic (GB)</th>
<th>Access-Time (10^3 hours)</th>
<th>C-Traffic</th>
<th>C-Time</th>
<th>W-Traffic</th>
<th>W-Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAME</td>
<td>1,227</td>
<td>3.87</td>
<td>15.15</td>
<td>13,669.71</td>
<td>0.38</td>
<td>2.98%</td>
<td>5.19%</td>
<td>0.76%</td>
<td>6.39%</td>
</tr>
<tr>
<td>NEWS_AND_READING</td>
<td>274</td>
<td>1.17</td>
<td>1.97</td>
<td>13,143.17</td>
<td>0.23</td>
<td>3.11%</td>
<td>2.91%</td>
<td>0.72%</td>
<td>3.95%</td>
</tr>
<tr>
<td>VIDEO</td>
<td>748</td>
<td>3.56</td>
<td>6.52</td>
<td>116,305.71</td>
<td>6.45</td>
<td>25.81%</td>
<td>1.22%</td>
<td>1.09%</td>
<td>10.91%</td>
</tr>
<tr>
<td>TOOL</td>
<td>227</td>
<td>3.84</td>
<td>9.43</td>
<td>77,329.87</td>
<td>0.68</td>
<td>15.63%</td>
<td>10.79%</td>
<td>4.40%</td>
<td>9.46%</td>
</tr>
<tr>
<td>SYSTEM_TOOL</td>
<td>217</td>
<td>3.37</td>
<td>7.54</td>
<td>34,012.16</td>
<td>0.25</td>
<td>3.05%</td>
<td>3.37%</td>
<td>2.17%</td>
<td>4.24%</td>
</tr>
<tr>
<td>SOCIAL</td>
<td>188</td>
<td>2.18</td>
<td>4.01</td>
<td>35,926.76</td>
<td>0.35</td>
<td>8.96%</td>
<td>4.77%</td>
<td>1.94%</td>
<td>5.66%</td>
</tr>
<tr>
<td>EDUCATION</td>
<td>172</td>
<td>1.68</td>
<td>2.98</td>
<td>13,893.55</td>
<td>0.34</td>
<td>1.46%</td>
<td>5.35%</td>
<td>0.87%</td>
<td>4.71%</td>
</tr>
<tr>
<td>LIFESTYLE</td>
<td>156</td>
<td>1.68</td>
<td>2.85</td>
<td>2,388.59</td>
<td>0.07</td>
<td>0.72%</td>
<td>1.00%</td>
<td>0.12%</td>
<td>1.06%</td>
</tr>
<tr>
<td>TRAVEL</td>
<td>111</td>
<td>1.62</td>
<td>2.75</td>
<td>8,182.24</td>
<td>0.03</td>
<td>0.78%</td>
<td>0.53%</td>
<td>0.52%</td>
<td>0.26%</td>
</tr>
<tr>
<td>PERSONALIZATION</td>
<td>104</td>
<td>1.49</td>
<td>3.68</td>
<td>7,420.38</td>
<td>0.86</td>
<td>0.85%</td>
<td>12.00%</td>
<td>0.46%</td>
<td>13.67%</td>
</tr>
<tr>
<td>FINANCE</td>
<td>90</td>
<td>0.52</td>
<td>0.50</td>
<td>382.60</td>
<td>0.02</td>
<td>0.13%</td>
<td>0.24%</td>
<td>0.02%</td>
<td>0.26%</td>
</tr>
<tr>
<td>COMMUNICATION</td>
<td>85</td>
<td>4.09</td>
<td>8.45</td>
<td>54,394.71</td>
<td>2.85</td>
<td>24.74%</td>
<td>49.01%</td>
<td>2.76%</td>
<td>35.26%</td>
</tr>
<tr>
<td>SHOPPING</td>
<td>78</td>
<td>1.57</td>
<td>3.00</td>
<td>21,808.51</td>
<td>0.07</td>
<td>3.16%</td>
<td>0.65%</td>
<td>1.32%</td>
<td>1.60%</td>
</tr>
<tr>
<td>PRODUCTIVITY</td>
<td>75</td>
<td>0.76</td>
<td>1.17</td>
<td>2,712.50</td>
<td>0.01</td>
<td>0.18%</td>
<td>0.17%</td>
<td>0.18%</td>
<td>0.26%</td>
</tr>
<tr>
<td>MOTHER_AND_BABY</td>
<td>48</td>
<td>0.10</td>
<td>0.15</td>
<td>525.72</td>
<td>0.01</td>
<td>0.07%</td>
<td>0.04%</td>
<td>0.03%</td>
<td>0.12%</td>
</tr>
<tr>
<td>MUSIC</td>
<td>43</td>
<td>2.33</td>
<td>3.39</td>
<td>49,540.12</td>
<td>0.17</td>
<td>5.66%</td>
<td>2.47%</td>
<td>3.08%</td>
<td>2.49%</td>
</tr>
<tr>
<td>SPORTS</td>
<td>27</td>
<td>0.51</td>
<td>0.36</td>
<td>61.49</td>
<td>0.00</td>
<td>0.02%</td>
<td>0.06%</td>
<td>0.00%</td>
<td>0.04%</td>
</tr>
<tr>
<td>IMAGE</td>
<td>23</td>
<td>0.14</td>
<td>0.17</td>
<td>80.64</td>
<td>0.00</td>
<td>0.06%</td>
<td>0.01%</td>
<td>0.05%</td>
<td>0.03%</td>
</tr>
<tr>
<td>TRAFFIC</td>
<td>14</td>
<td>0.10</td>
<td>0.12</td>
<td>78.10</td>
<td>0.00</td>
<td>0.02%</td>
<td>0.03%</td>
<td>0.00%</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

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

*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

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

Figure 12: Network Activity Distribution among User Groups
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