**ASONAM 2019**
**Industrial Track**

# Uncovering **Download Fraud** Activities in Mobile App Markets

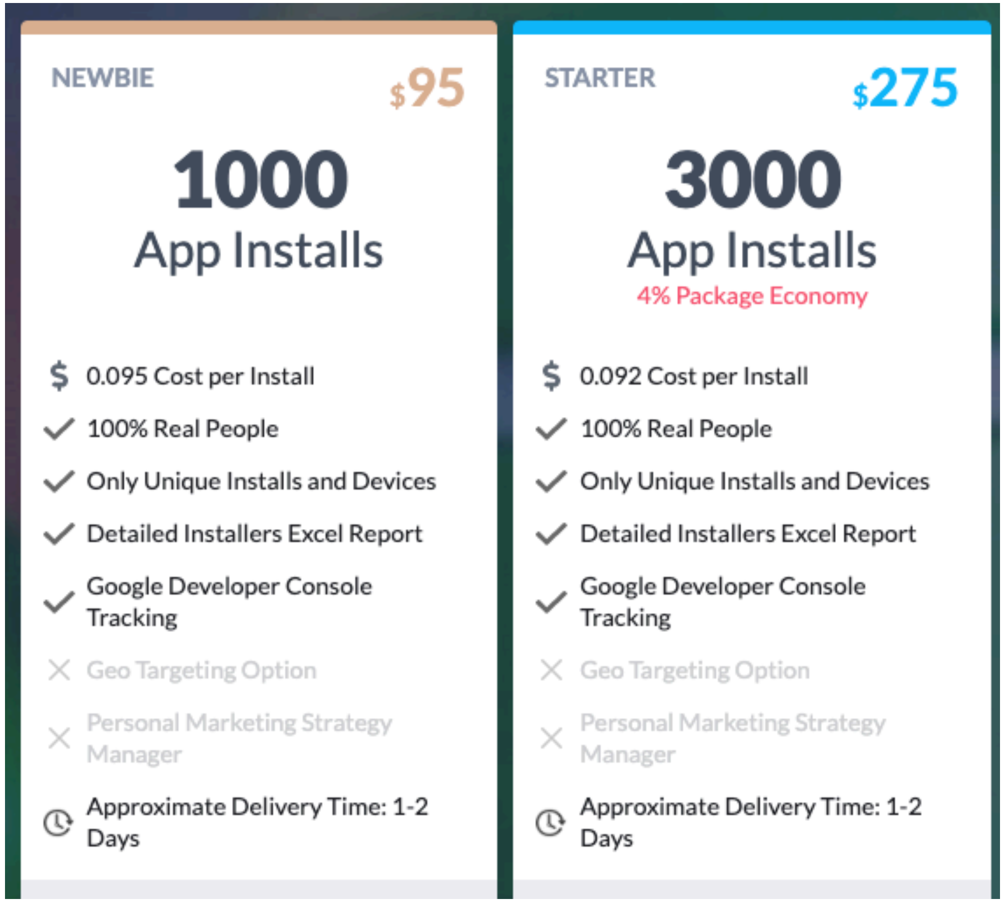
Yingtong Dou, Weijian Li, Zhirong Liu,
Zhenhua Dong, Jiebo Luo, Philip S. Yu

**ydou5@uic.edu**

**Slides are available at http://ytongdou.com/files/asonam19slides.pdf**

Aug. 28, 2019

# Fake Downloads are Prevalent



NEWBIE $95

**1000**
App Installs

- $ 0.095 Cost per Install
- ✓ 100% Real People
- ✓ Only Unique Installs and Devices
- ✓ Detailed Installers Excel Report
- ✓ Google Developer Console Tracking
- ✗ Geo Targeting Option
- ✗ Personal Marketing Strategy Manager
- ◷ Approximate Delivery Time: 1-2 Days

STARTER $275

**3000**
App Installs
4% Package Economy

- $ 0.092 Cost per Install
- ✓ 100% Real People
- ✓ Only Unique Installs and Devices
- ✓ Detailed Installers Excel Report
- ✓ Google Developer Console Tracking
- ✗ Geo Targeting Option
- ✗ Personal Marketing Strategy Manager
- ◷ Approximate Delivery Time: 1-2 Days

Background

# Threats & Challenges

- Threats
  - 10% downloads/installs in mobile App markets are fake which cost near $300 million loss in marketing in 2018[1]
  - Fake downloads mislead the recommender system and advertisement bidding system

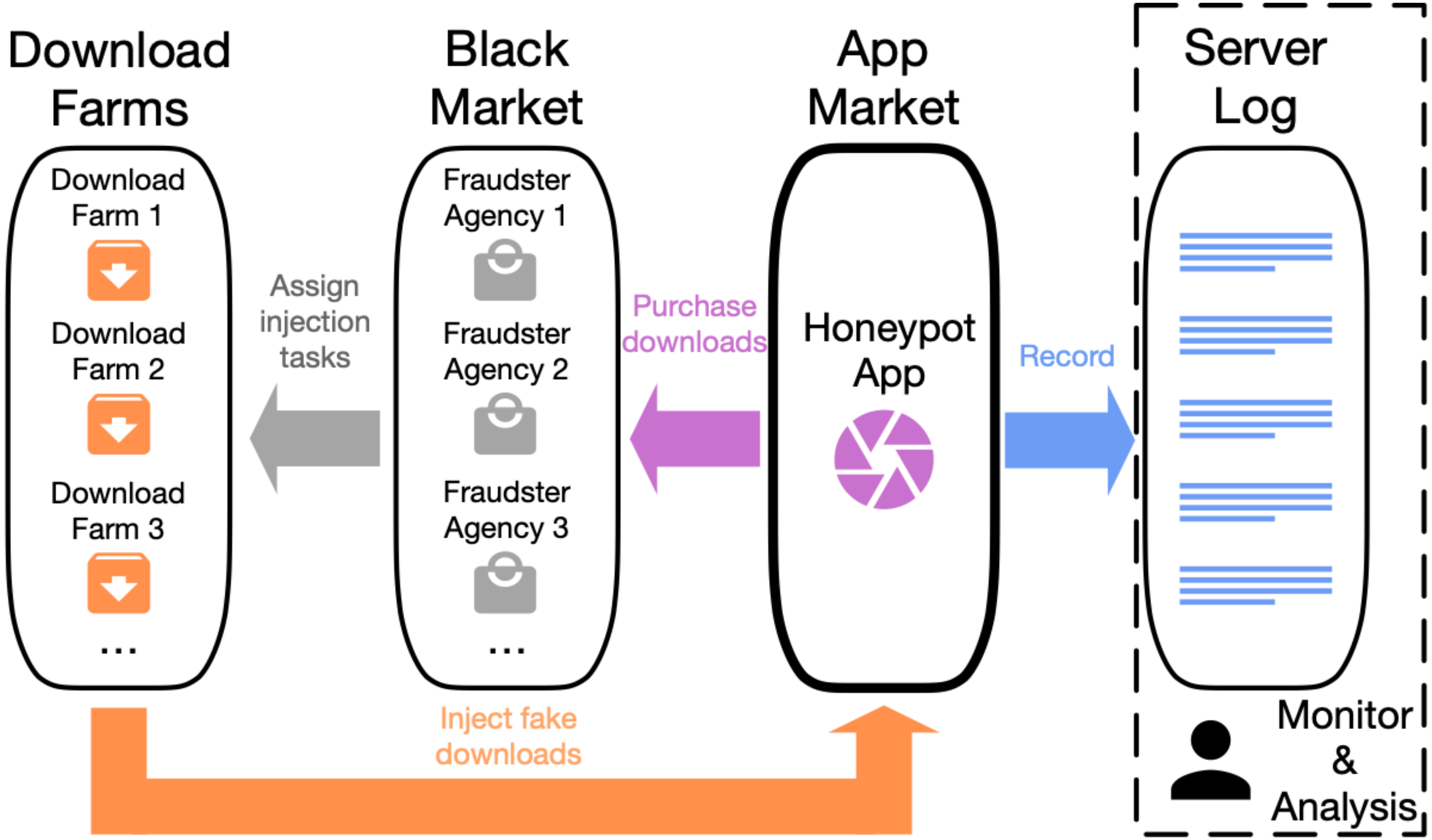
- Challenges
  - Mixed multi-source fake downloads
  - Lack of ground truth

[1]. "The underworld of app install advertising," Datavisor Threat Labs, https://www.datavisor.com/blog/datavisor-threat-labs-report-the-underworld-of-app-install-advertising/

# Research Questions

- RQ1: What are the types of download fraud activities in the App market?
- RQ2: How to identify the download fraud activities?
- RQ3: How to mitigate the download fraud in App markets?

# Setting Up the Honeypot



Background Honeypot

# Download Fraud Types

- Type I: Boosting Front End Downloads
  - Like click fraud in online advertisement
  - Employ automated scripts to inject fake clicks
  - A prevalent attack with low budget
  - Minor threat to App markets backbone

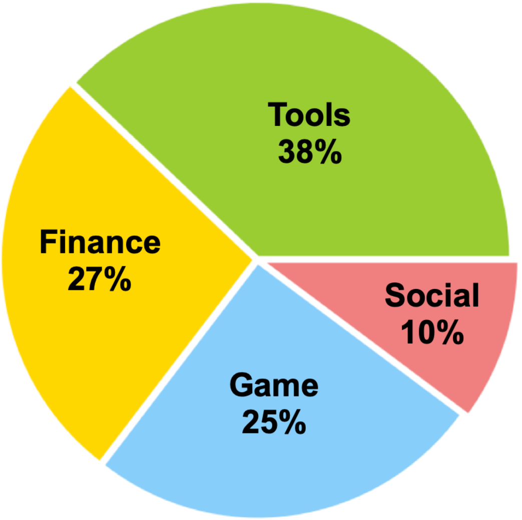
**All fake downloads injected to the honeypot fall into this category**

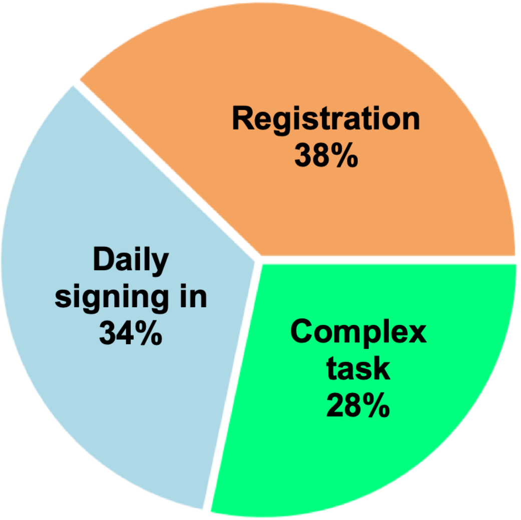
# Download Fraud Types

- Type II: Optimizing App Search Ranking
  - Biasing search/recommendation algorithms via imitating real devices search/download/install behavior
  - Usually launch with App Store Optimization (ASO)
  - Medium budget, high threat, hard to detect

# Download Fraud Types

- Type III: Enhancing User Acquisition & Retention Rate
  - Complex tasks implemented by crowd workers
  - High budget, very hard to be detected
  - Low threat to App markets
  - Cheat venture capital and advertiser





# Identifying Fake Downloads (Type I)

- Type I: Determine and filter fake downloads by **Source** and **Device** information

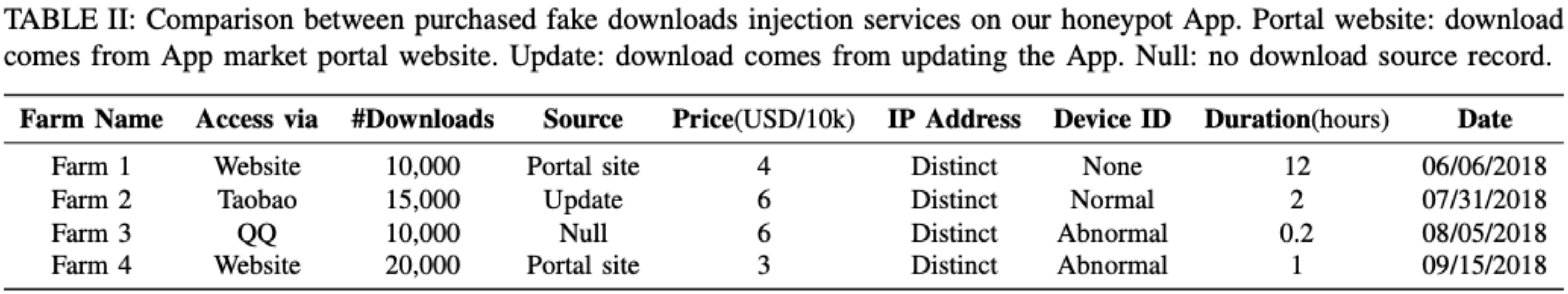
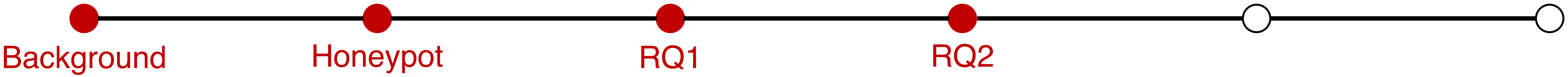


TABLE II: Comparison between purchased fake downloads injection services on our honeypot App. Portal website: download comes from App market portal website. Update: download comes from updating the App. Null: no download source record.

| **Farm Name** | **Access via** | **#Downloads** | **Source** | **Price**(USD/10k) | **IP Address** | **Device ID** | **Duration**(hours) | **Date** |
|---|---|---|---|---|---|---|---|---|
| Farm 1 | Website | 10,000 | Portal site | 4 | Distinct | None | 12 | 06/06/2018 |
| Farm 2 | Taobao | 15,000 | Update | 6 | Distinct | Normal | 2 | 07/31/2018 |
| Farm 3 | QQ | 10,000 | Null | 6 | Distinct | Abnormal | 0.2 | 08/05/2018 |
| Farm 4 | Website | 20,000 | Portal site | 3 | Distinct | Abnormal | 1 | 09/15/2018 |



Background

Honeypot

RQ1

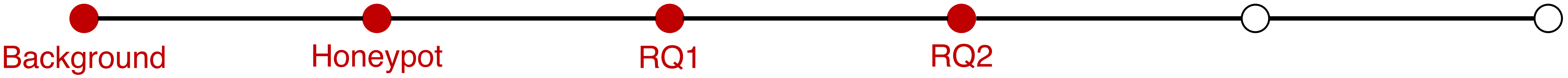
RQ2

# Identifying Fake Downloads (Type II)

- Ground Truth
  - **Positive Downloads:** All downloads from Apps where half of the downloads from non-vendor devices
  - **Negative Downloads:** Downloads from vendor-verified devices
- Data Collection
  - Dataset sampled from an Android App Market download logs during May 2018 to December 2018
  - One million positive samples, nine million negative samples
  - Logs include no privacy information, all IDs are secured by hashing

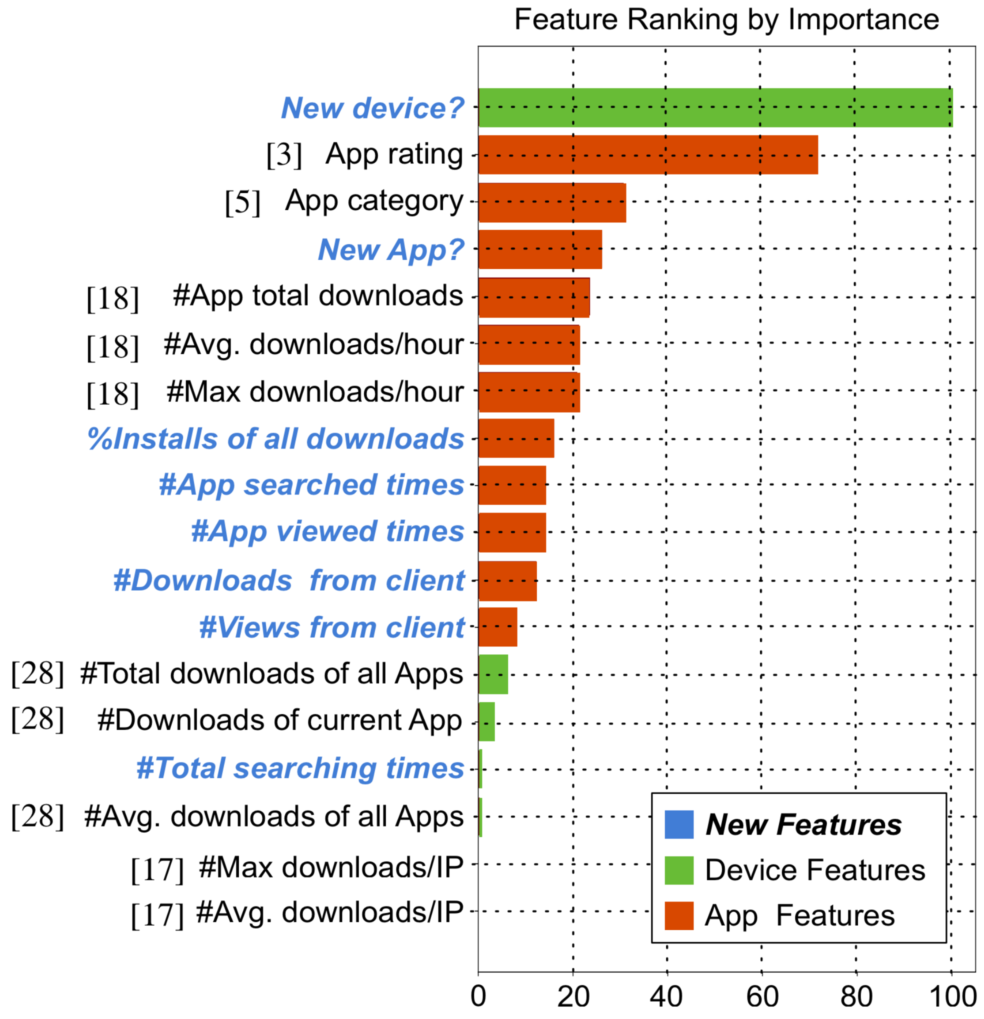
# Identifying Fake Downloads (Type II)

- Feature Selection (***New features***)
  - Device features:
    - ***New device?***; Total downloads from all Apps;
    - Downloads from current App; Avg. downloads of all Apps;
    - ***Total searching times***; Max. downloads/IP; Avg. downloads/IP
  - App features:
    - App rating; App category; ***New App?***; App total downloads;
    - Avg. downloads/hour; Max. downloads/hour; ***%Installs***;
    - ***App searched times***; ***App viewed times***; ***Downloads from client***; ***Views from client***;



# Identifying Fake Downloads (Type II)

- Feature Importance
  - All features are extracted from a download record
  - Calculated by Gini Impurity using Random Forest
  - Categorical features are processed with one-hot encoding



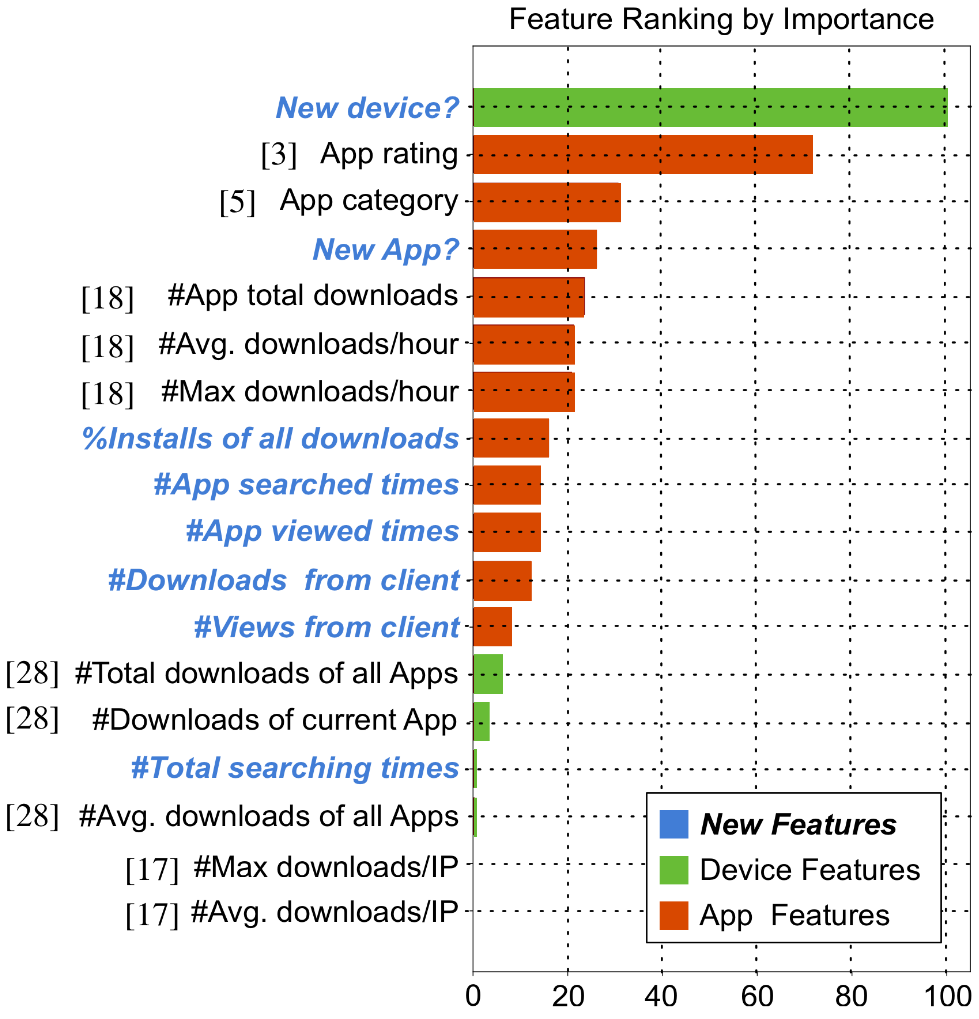
# Identifying Fake Downloads (Type II)

- Observations

  - **New device?** indicates a download bots reset their device IDs

  - **New App?** indicates many Apps soliciting fake downloads are newly released

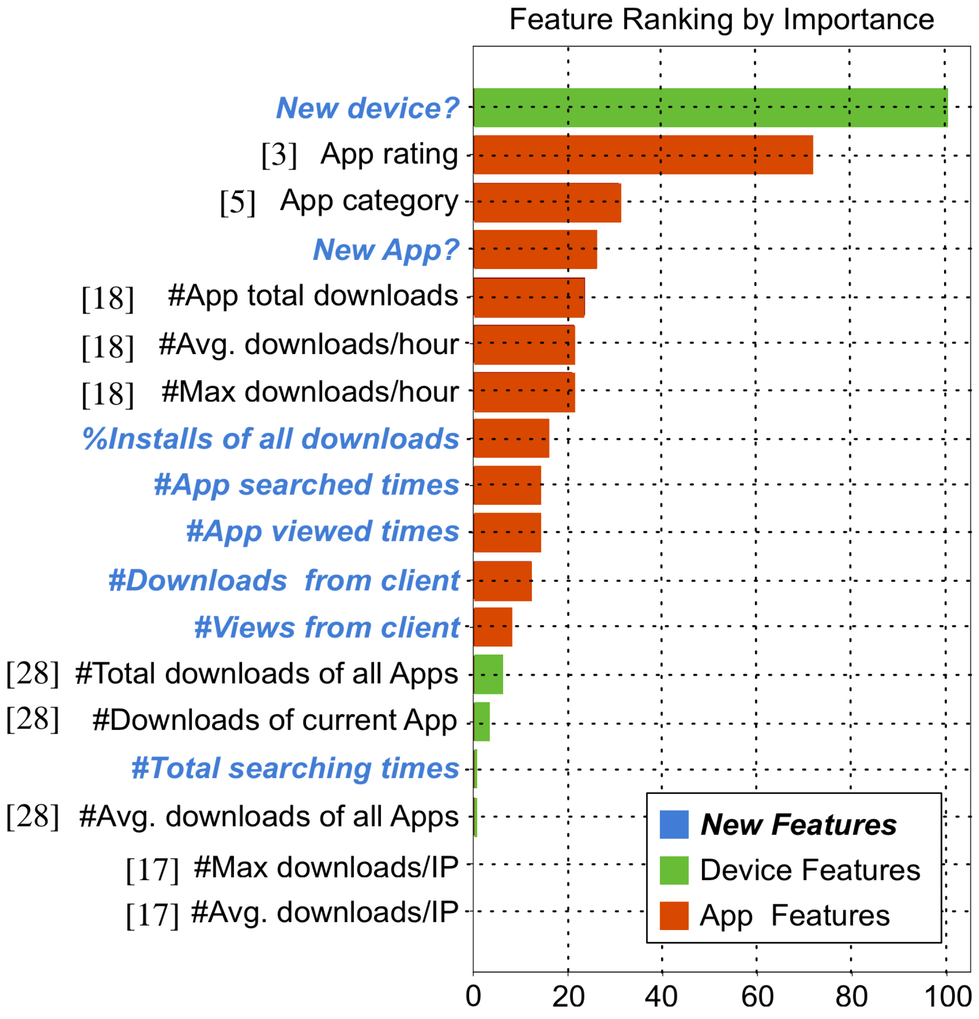
  - App statistical features are useful in distinguishing fake downloads


Feature Ranking by Importance

Background Honeypot RQ1 RQ2

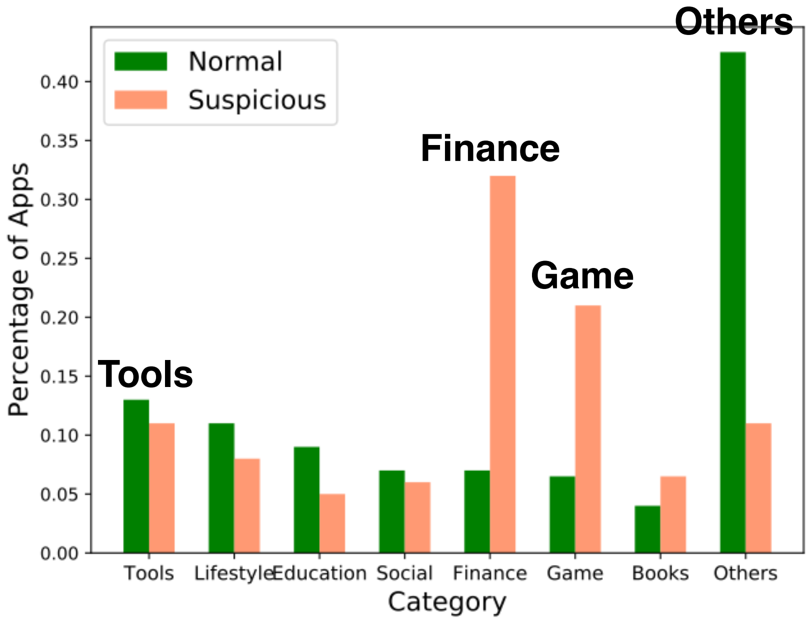
# Identifying Fake Downloads (Type II)

- Observations (cont'd)
  - Except the **New device?** feature, most App features are more useful than device features
  - Behavioral features and IP-based features are useless, illustrating that the bots could imitate regular user behavior

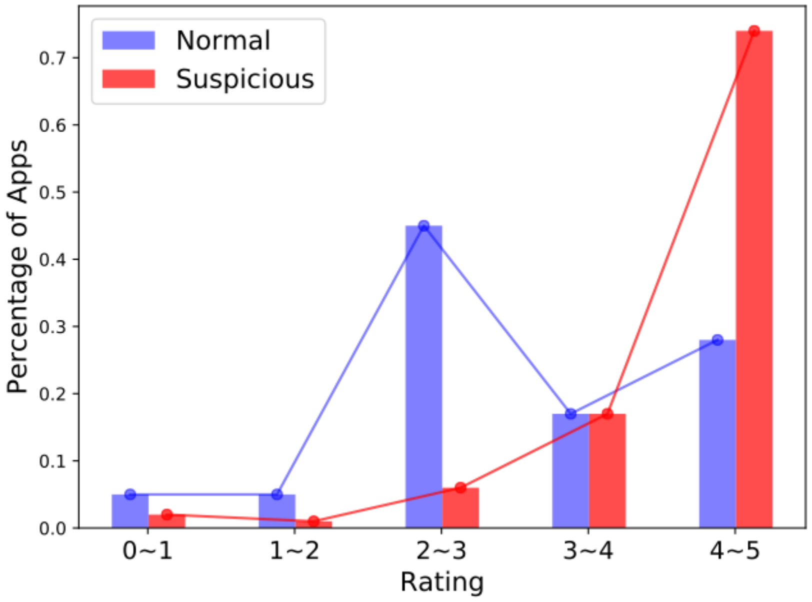


# Identifying Fake Downloads (Type II)

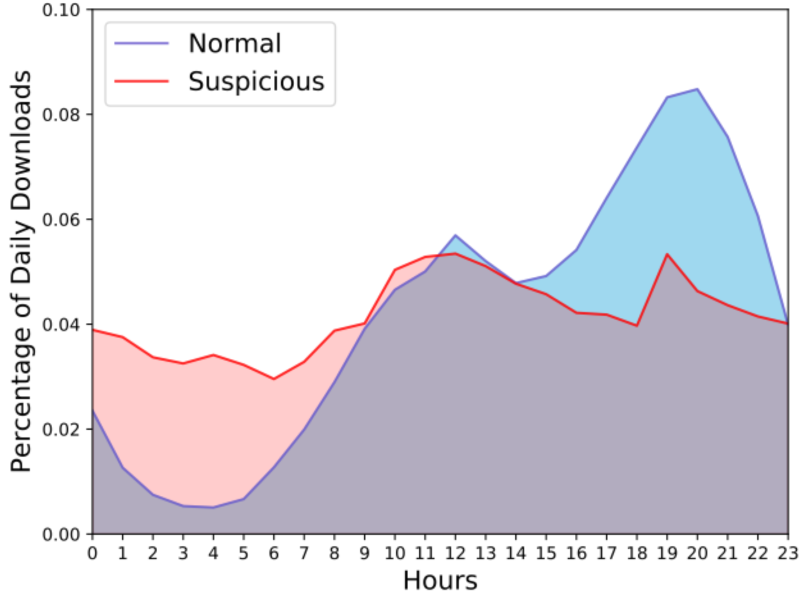
- Comparative analysis



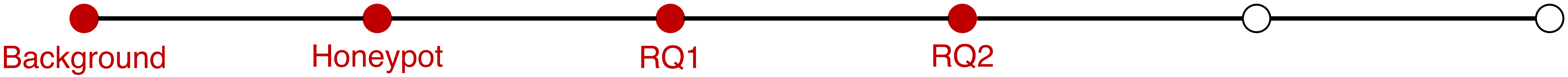
Category Distribution



Rating Distribution



Traffic in A Day



Background

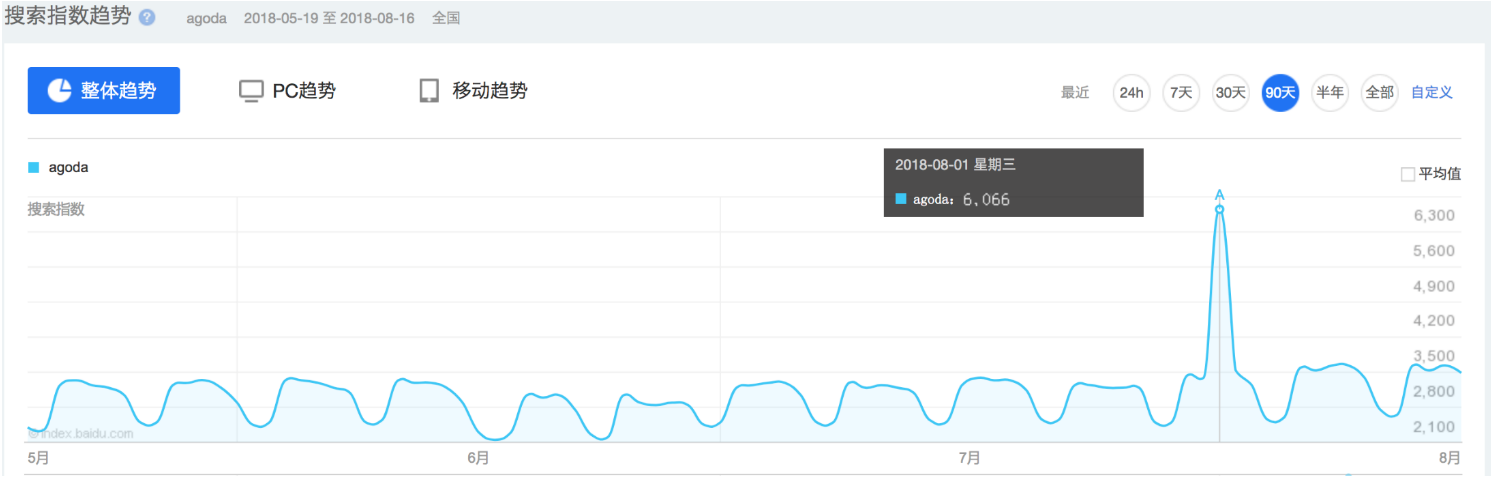
Honeypot

RQ1

RQ2

# Identifying Fake Downloads (Type II)

- Two extra points
  - Not all anomalies are suspicious



搜索指数趋势 agoda 2018-05-19 至 2018-08-16 全国

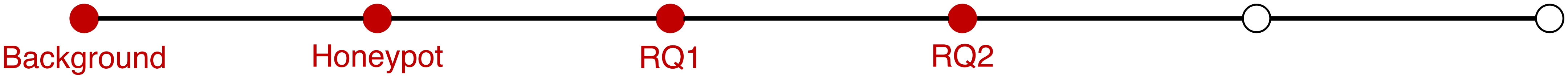
整体趋势 PC趋势 移动趋势

最近 24h 7天 30天 90天 半年 全部 自定义

2018-08-01 星期三

agoda：6,066

  - Download fraud traffic has a correlation with trending events



# Stances from Three Parties

- App Marketer
  - Sometimes fake downloads are cheaper than regular advertisement, and injected fake downloads could help meet the KPI
- Fraudster Agency
  - Most fraudster agencies are disguised as marketing firms, fake downloads injection is part of the ASO bundle
  - Long-term cooperation between App operators and fraudster agencies are prevalent, especially the Gaming Apps.
- Market Operator
  - Fake downloads are not 100% negative for App markets. They could facilitate App releasings which always face cold start problems

# How to Mitigate Download Fraud?

- Adapting the agility of fraudsters
- Building suspicious behavior signature database
- Crafting diversified anti-fraud mechanism
- Devising fine-grained advertisement services
- Elaborating clear incentives and sanctions

# Key Takeaways

- Fake downloads are generated from multiple channels which have different goals

- Rule-based algorithm usually has a high false-positive rate. We need integrate information from multiple sources to justify suspiciousness

- Attracting marketers to legitimate promotion channels is more important than filtering fake downloads

# Thank you!
# Q & A

Yingtong Dou @ University of Illinois at Chicago

ydou5@uic.edu

Slides are available at http://ytongdou.com/files/asonam19slides.pdf