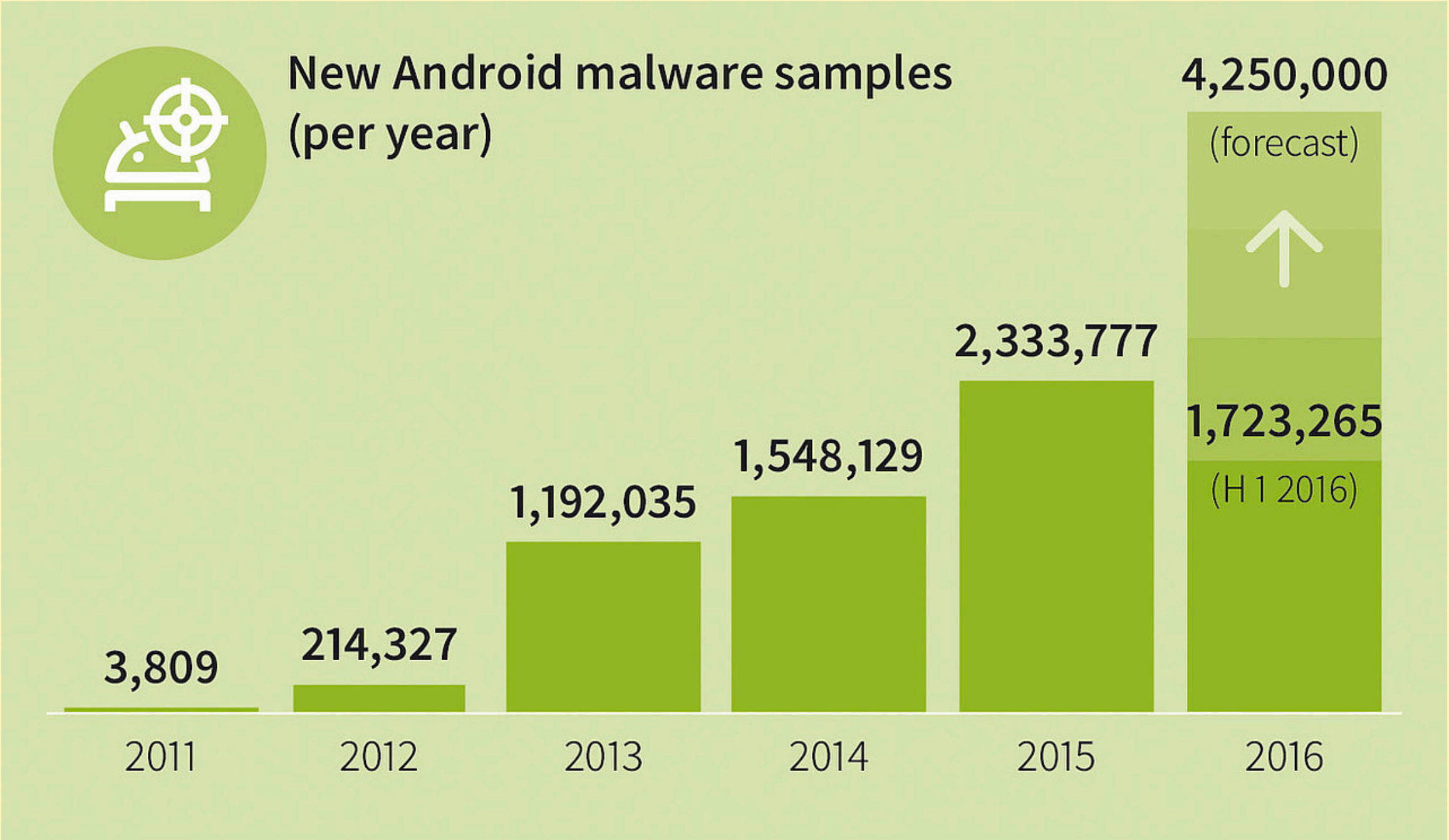
Non-intrusive Detection of Mobile Malware and Botnets

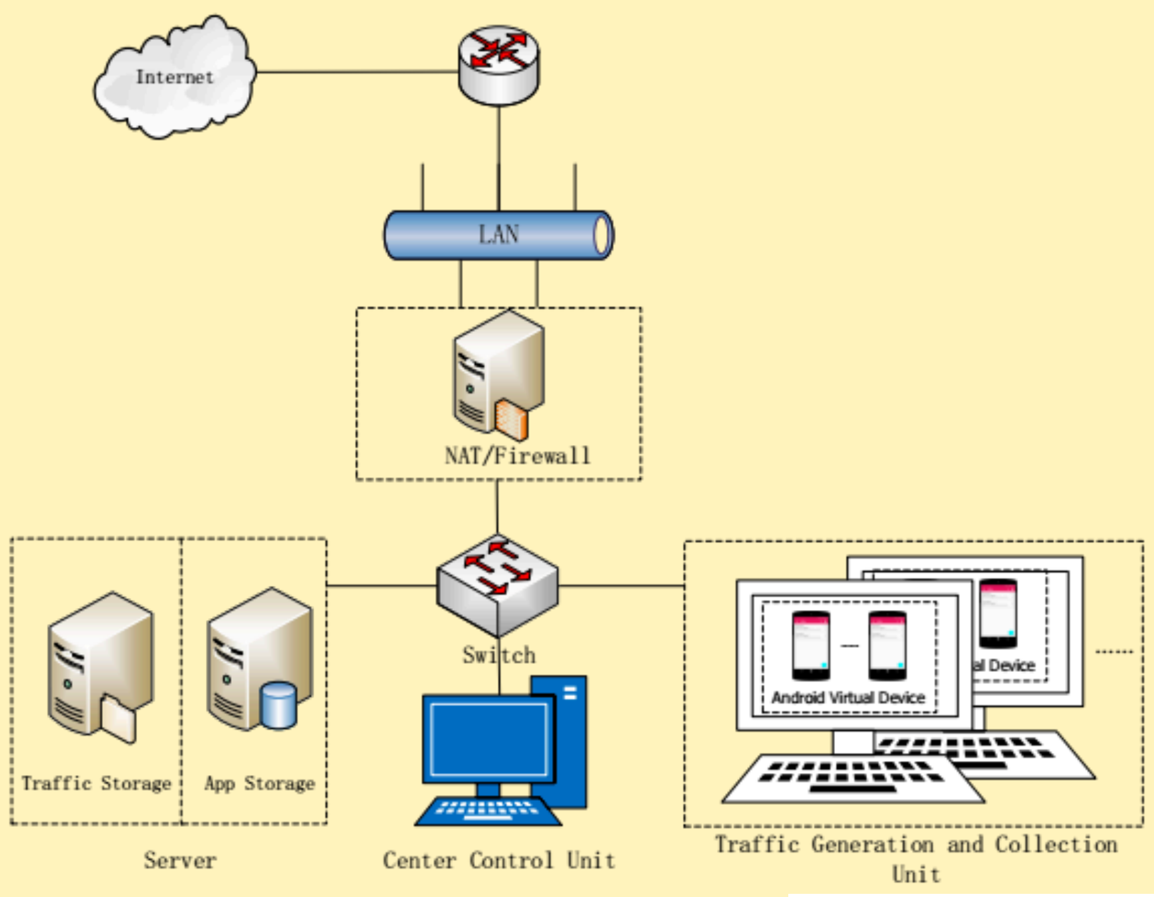
PI: Qiben Yan

Project URL: Http://think.unl.edu/crii-research.html

NSF Award #: 1566388

The objective of this project is to develop technologies that will detect mobile malware's **malicious network activity** at the gateway of a large-scale network, and mitigate the network-wide damage or harm that might be caused by malware apps operating inappropriately or maliciously.





Design of Application Traffic Generator

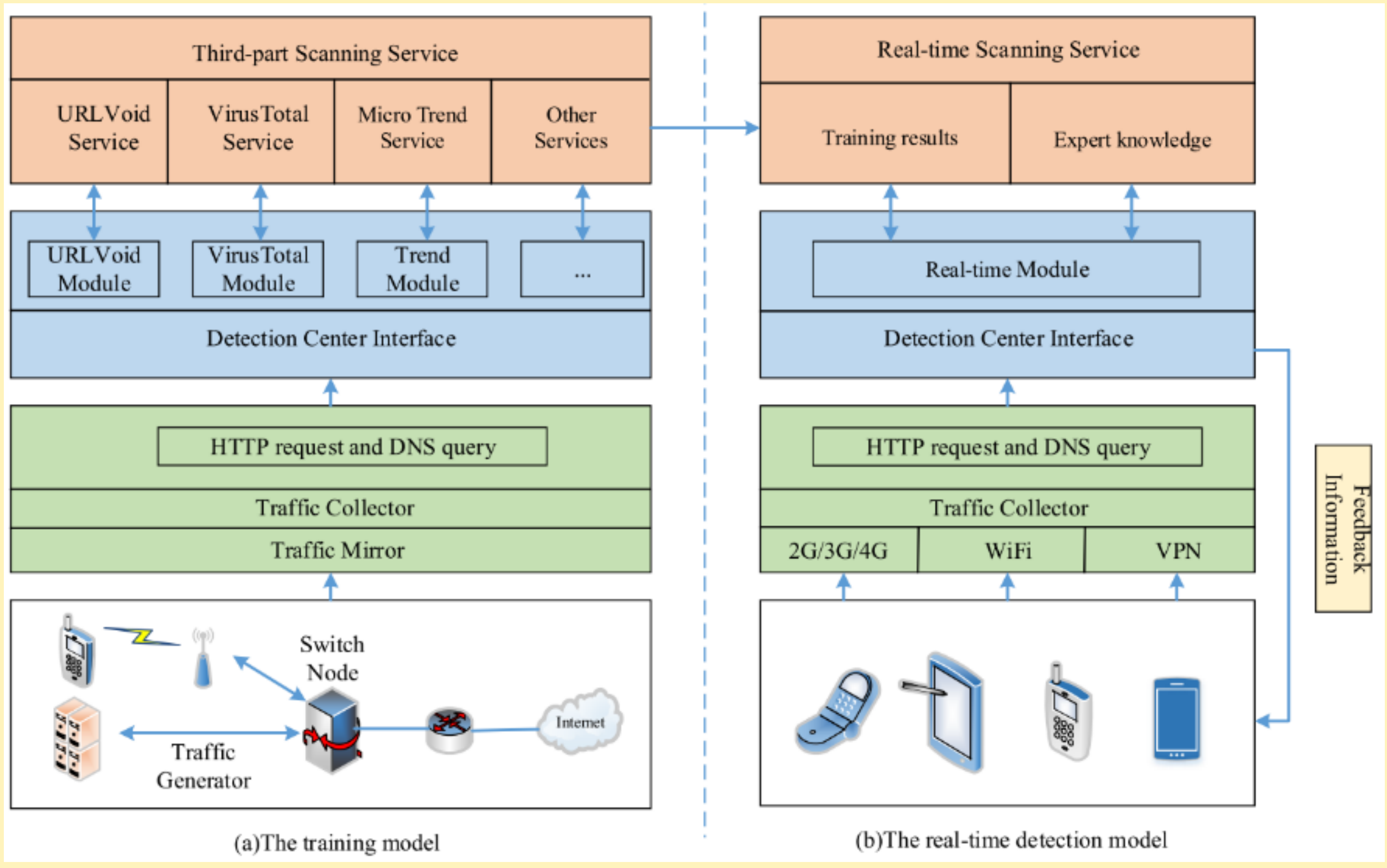
- •Automated traffic generation
- •High performance and scalable framework
- •High quality application traffic dataset

Background

- Android allows to install applications from uncertified third party stores
- 97% of all mobile malicious applications target Android[1]
- A new Android malware appears every 11 seconds[2]

[1] Forbes Tech, http://www.forbes.com/sites/gordonkelly/2014/03/24/report-97-of-mobile-malware-is-on-android-this-is-the-easy-way-you-stay-safe/#3784dff87d53, 2014

[2]GDATA MOBILE MALWARE REPORT, https://file.gdatasoftware.com/01_public/Presse/Publikationen/Malware_Reports/EN/G_DATA_MobileMWR_Q4_2015_EN.pdf, 2015



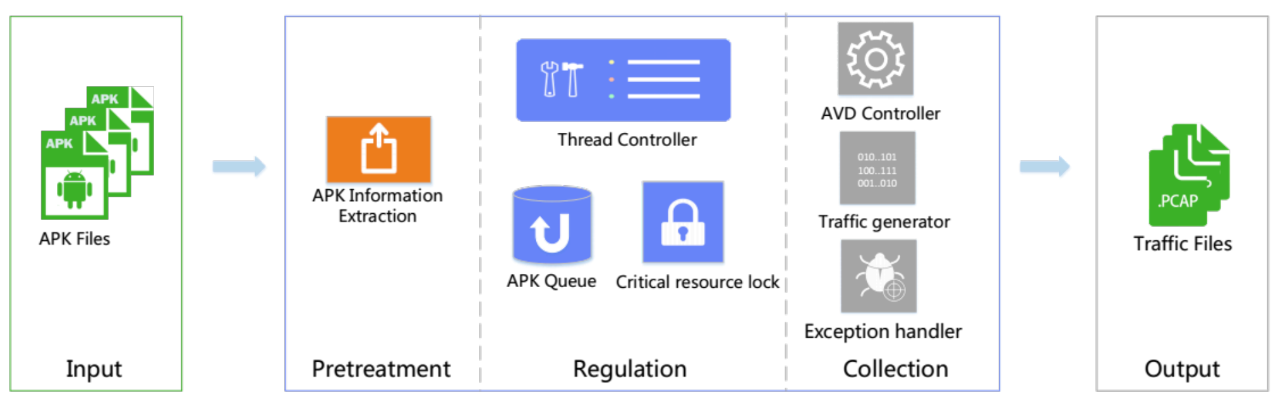
(a)The training model

(b)The real-time detection model

Approach

- **Mobile malware traffic collection**: use program analysis to identify network-related APIs, and to develop triggering mechanisms
  - Identify the HTTP API and corresponding execution path
  - Develop static analysis tools to discover those suspicious HTTP APIs and extract the API call graph
  - Design effective inputs to activate the call graph, which in turn generates malicious network traffic for collection
- **P2P/HTTP botnet detection and mobile botnet characterization:** evaluate the aggregated network behavior from multiple interactive bots
- **Network-based mobile malware detection:** use data analytics to identify mobile malware in real time using application-layer traffic
  - Extract features related to program execution sequences and the lexical contexts from HTTP/DNS traffic, such as the key value pair information in the HTTP request
  - The extracted traffic features need to be robust and reliable enough to avoid being evaded by smart malware developers
  - The feature extraction and detection mechanism must be efficient enough to be deployed in real time
  - Investigate the evolution of mobile botnet and the relationship between mobile botnet and PC botnet

DroidCollector: Automated Malware Traffic Generator



DroidCollector:

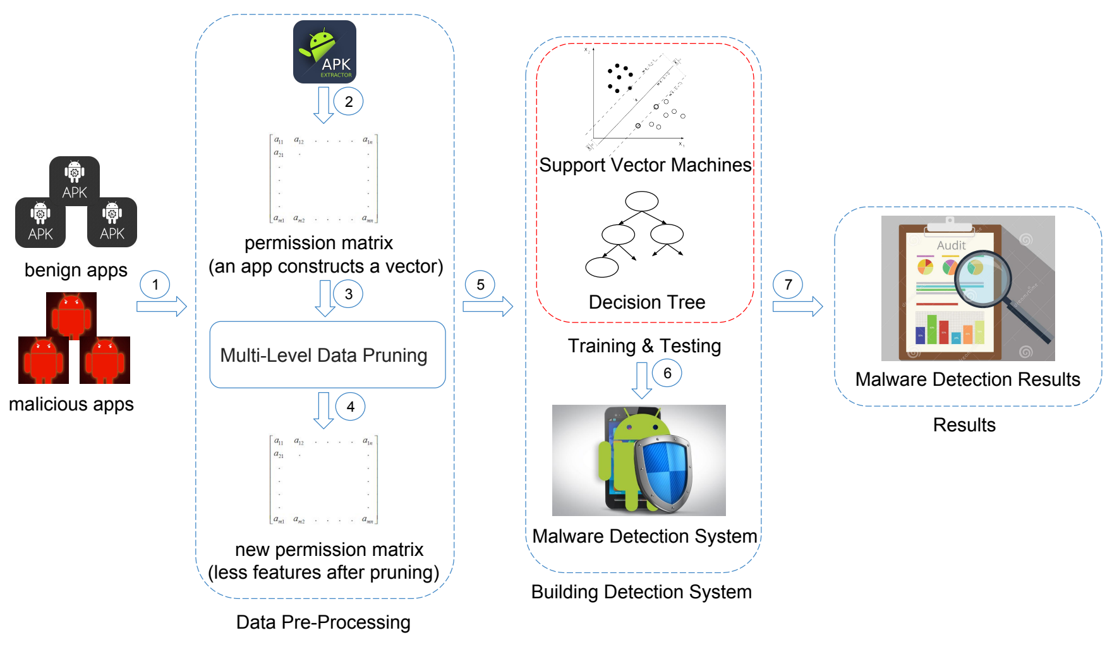
•Leverages multithreading to perform active and automatic network traffic collection

•Collects 808 MB and 330 MB traffic data generated by 6000 benign apps and 5560 malicious apps in a short period of time

SigPID: Significant Permission Identification for Android Malware Detection

SigPID:

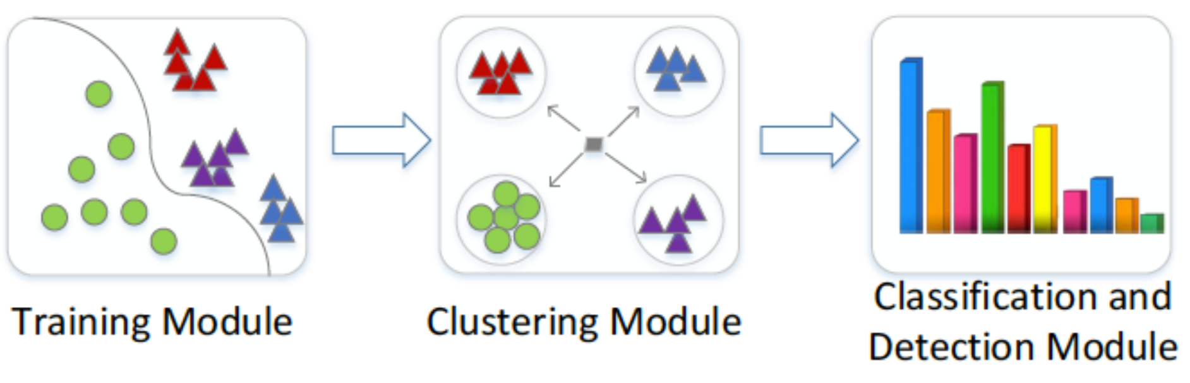
- Identify significant permissions for real-time malware detection
- Provide Multi-Level Data Pruning (MLDP)
- Perform malware detection using only significant permissions



DroidClassier: Adaptive Mining of Application-Layer Header

DroidClassifier:

- Multiple HTTP header fields as features
- A novel weighted score-based metric for malware classification
- Performance is optimized via both supervised and unsupervised learning

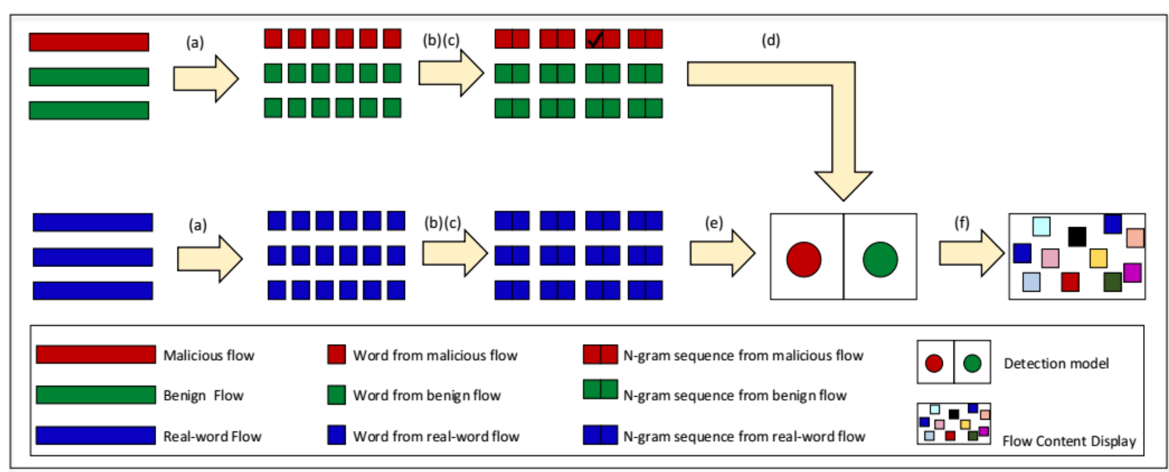


TextDroid: Semantics-based Detection of Mobile Malware Using Network Flows

TextDroid:

•HTTP flow headers are segmented into words, which are supplied to generate the bag-of-words using an N-gram generation method

•Automatically identifies and extracts the distinguishable features



PI: Qiben Yan, Dept. of Computer Science and Engineering, University of Nebraska Lincoln, Lincoln, Nebraska, Email: yan@unl.edu, Phone: 402-472-5075

NSF Secure and Trustworthy Cyberspace Inaugural Principal Investigator Meeting

National Science Foundation
WHERE DISCOVERIES BEGIN