ACSAC 2023 • December 4-8, 2023 • Austin, Texas, USA

ANDetect: A Third-party Ad Network Libraries Detection Framework for Android Applications

Xinyu Liu[1,2], Ze Jin[1,2], Jiaxi Liu[1,2], Wei Liu[1,2], Xiaoxi Wang[1,2], Qixu Liu[1,2]

[1]Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Cyber Security, University of Chinese Academy of Sciences

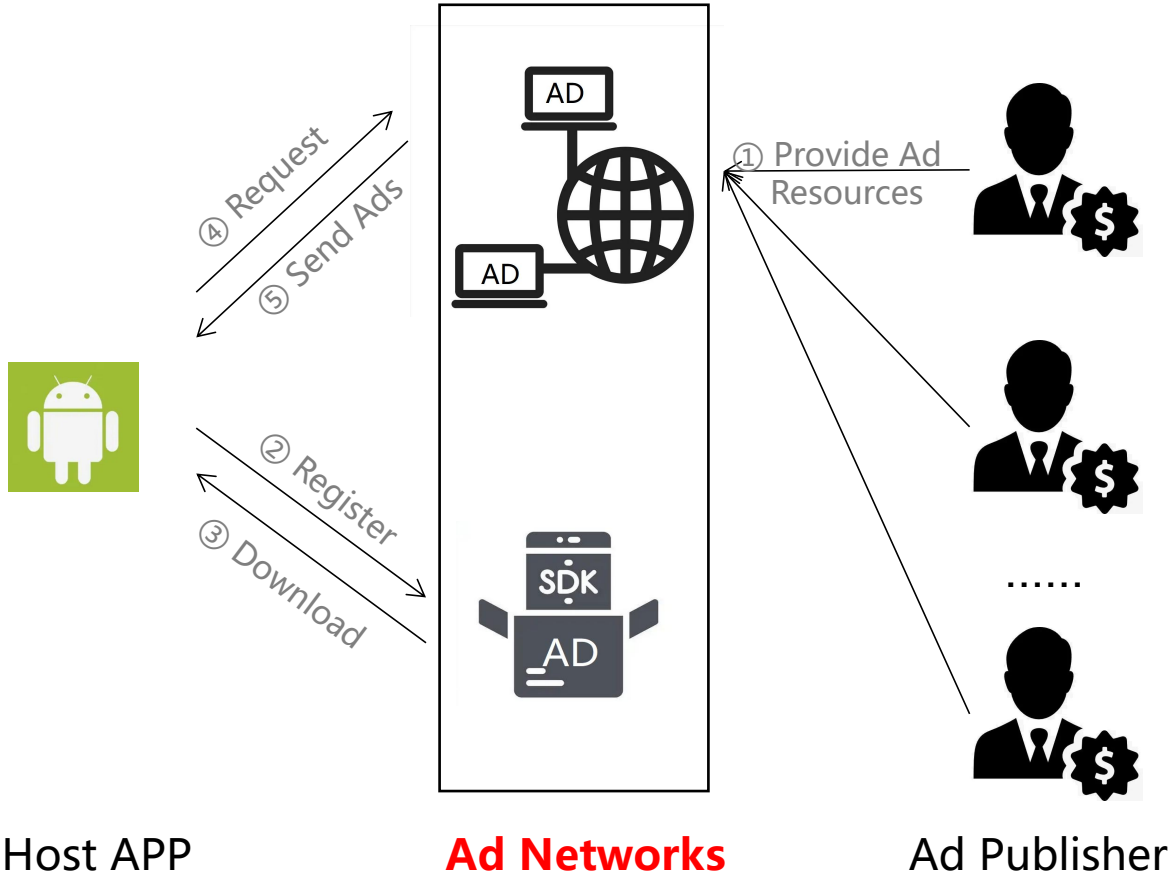
December 5, 2023

Artifacts Evaluated acm Functional V1.1

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

中国科学院大学
University of Chinese Academy of Sciences

Advertising Networks



Meta

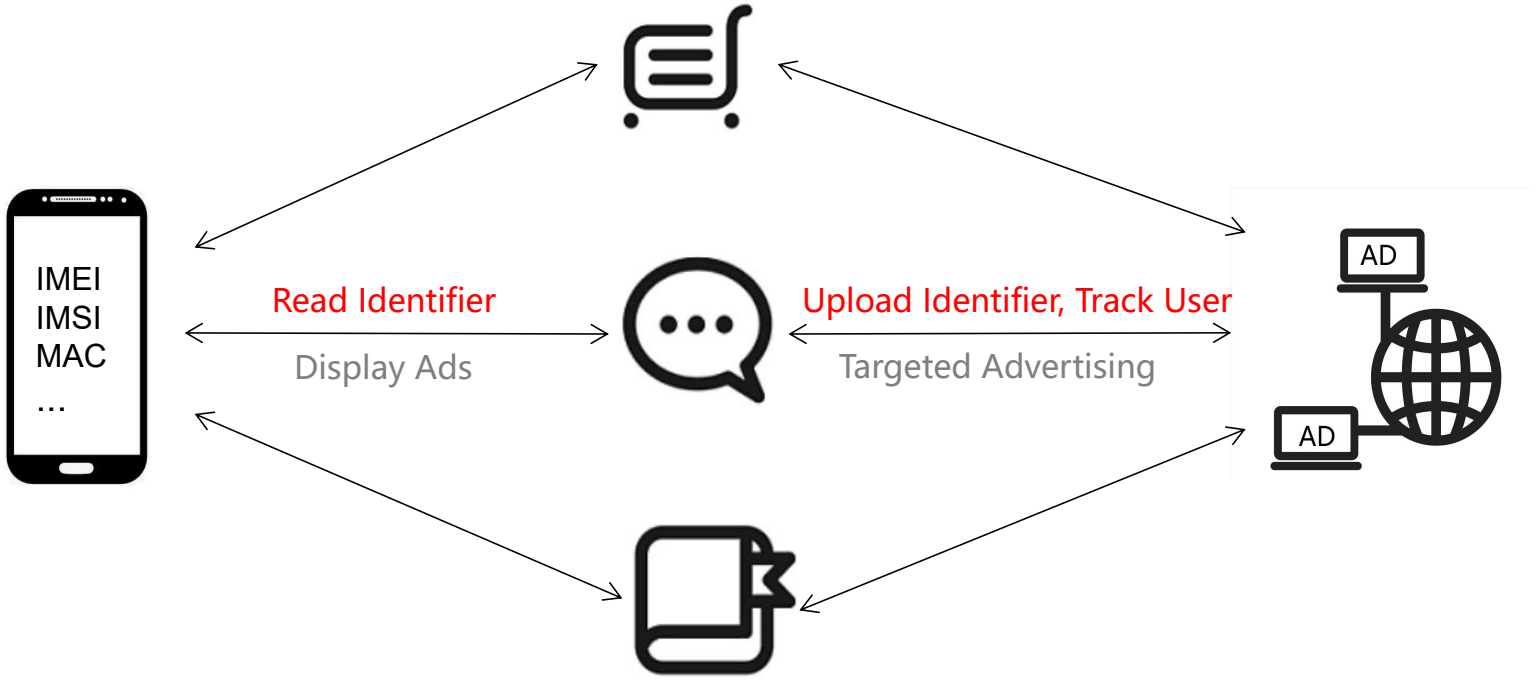
APPLOVIN

Google AdMob

Yandex

Why Detect Ad Network Libs?

A1: Protect user's privacy.



Tracking

⬇

Analyzing

⬇

User Portrait

...

Why Detect Ad Network Libs?

A2: Identify adware effectively.

Adware is a type of malicious software that secretly installs itself on your device and displays unwanted advertisements and pop-ups.

APP Obfuscation and Encryption

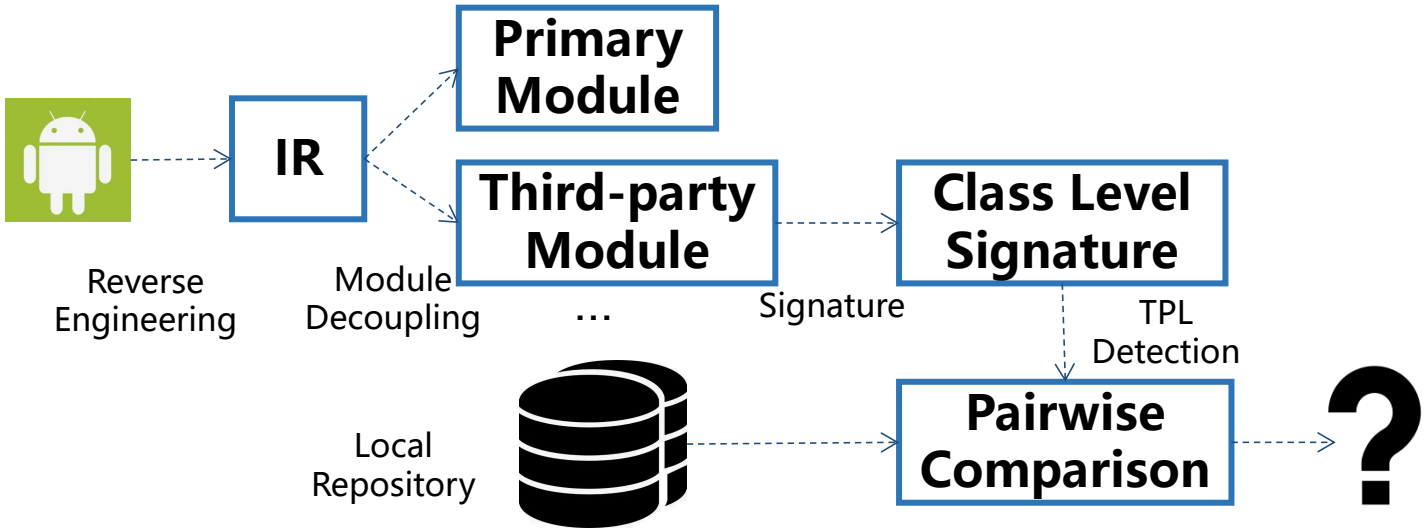
Identifier Renaming	APP Obfuscation	allatori, DashO, DexGuard, ProGuard
String Encryption		
Package Flattening		
Dead Code Removal		
Dex Encryption	APP Encryption	Qihoo, APKProtect

…

Existing Work

Recognize TPL

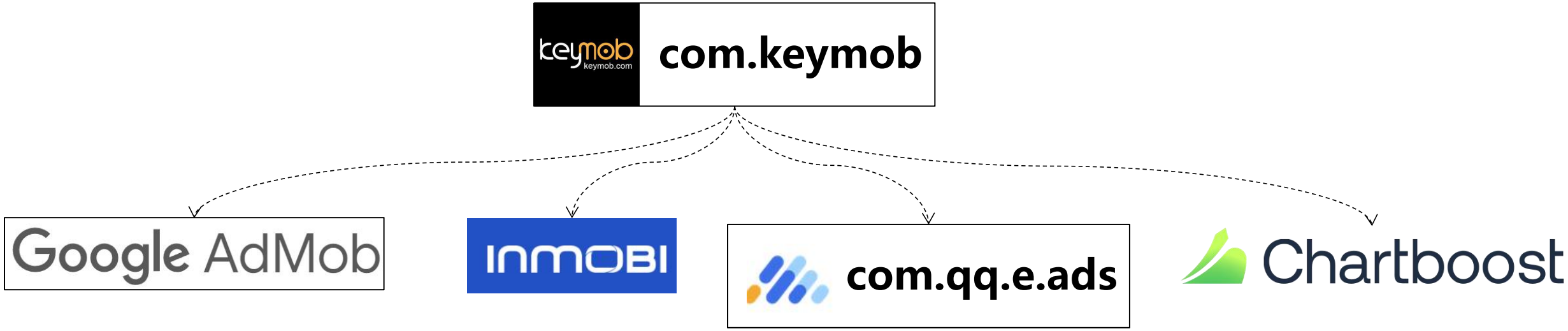
- LibRadar (ICSE’16)
- LibScout (CCS’16)
- LibD (ICSE’17)
- LibLoad (TMC’22)



- The essence of these works is whitelist-based method.
- None of these works can detect the ad libs in an encrypted APK.

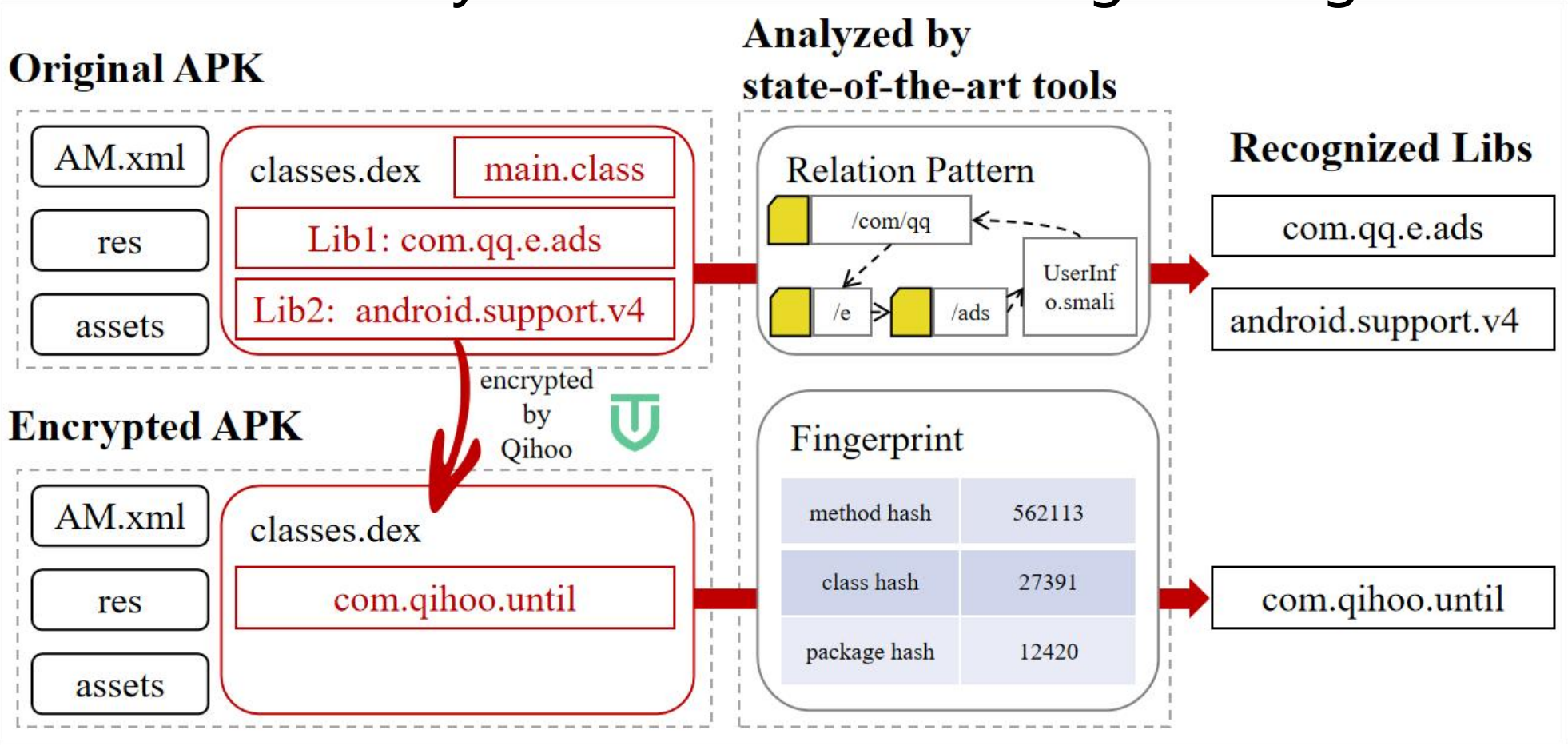
Challenges

C1: Some integrated ad network libraries may be incorporated by many single ad libraries. Traditional whitelist-based methods fail.



Challenges

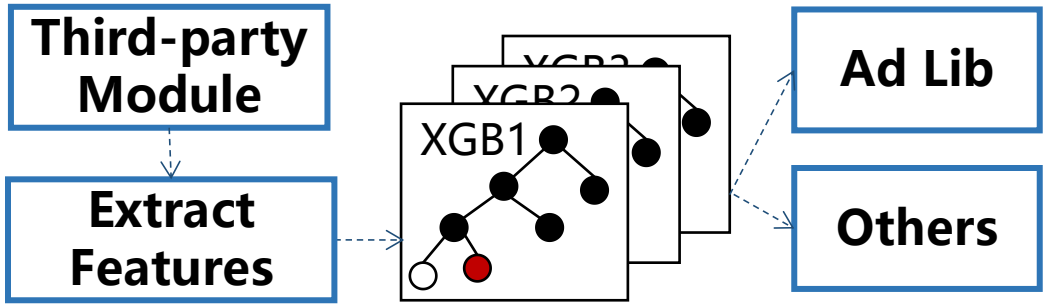
C2: The encryption of APP is always used by developers to enhance security and deter reverse engineering.

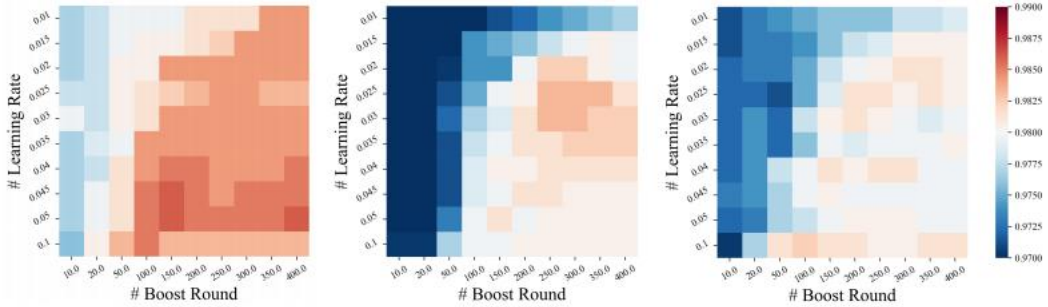


Break the Limitation of Whitelist

C1: Some integrated ad network libraries may be incorporated by many single ad libraries. Traditional whitelist-based methods fail.

- **Extract the features of system API.**
  - API count
  - Distribution across classes
  - Distribution across methods
- **Train 3 Xgboost models with Dataset_{AN} and Dataset_{NAN}.**
  - 833 ad libraries
  - 2758 non-ad libraries
- **Integrate the models and classify every third-party module.**





(a) API count (b) DIST across classes (c) DIST across methods

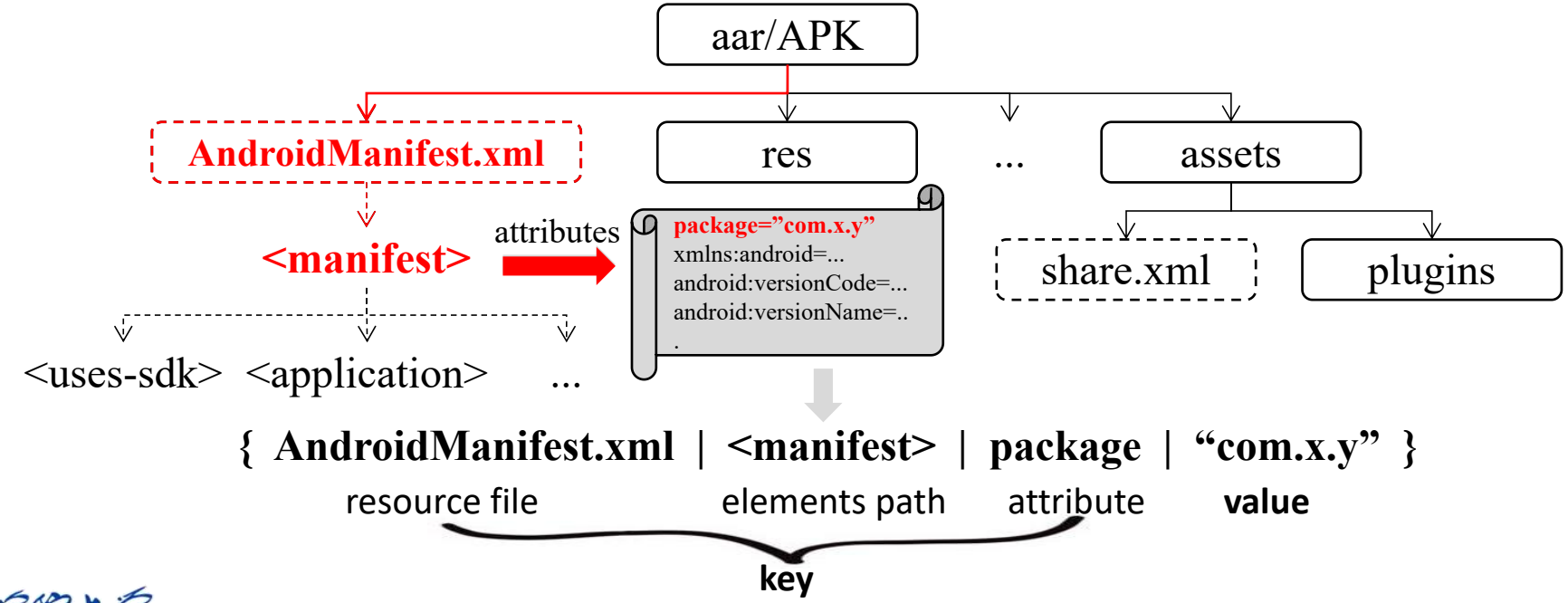
Address the Challenge of Encrypted APPs

C2: The encryption of APP is always used by developers to enhance security and deter reverse engineering.

- **Utilize the resource features of encrypted APPs.**

- **Resource Features**
  - AndroidManifest.xml
  - files in *res*
  - files in *assets*

- **Key-value Mapping**
  - resource file
  - elements path
  - attribute



Address the Challenge of Encrypted APPs

Q: How to choose suitable resource features and how to match?

- **Universality.**

$$U_k = \frac{\sum_{i=1}^{l} In_{k,K_i}}{l}, \quad In_{k,K_i} = \begin{cases} 1, & k \in K_i \\ 0, & k \notin K_i \end{cases}$$

- **Differentiation.**

$$d_{v,i} = \max_{e \in v}\{\frac{f_{e,i}}{\sum_{e' \in e_i} f_{e',i}} \times \ln \frac{l}{1 + \sum_{j=1}^{l} In_{e,e_j}}\}$$

$$D_k = \frac{\sum_{i=1}^{l} \sum_{v \in M_{k,i}} d_{v,i}}{\sum_{i=1}^{l} |M_{k,i}|}$$

- **Insensitivity.**

$$IS_k = \frac{\sum_{i=1}^{|an_k|} \frac{1}{|v_{k,i}|}}{|an_k|}$$

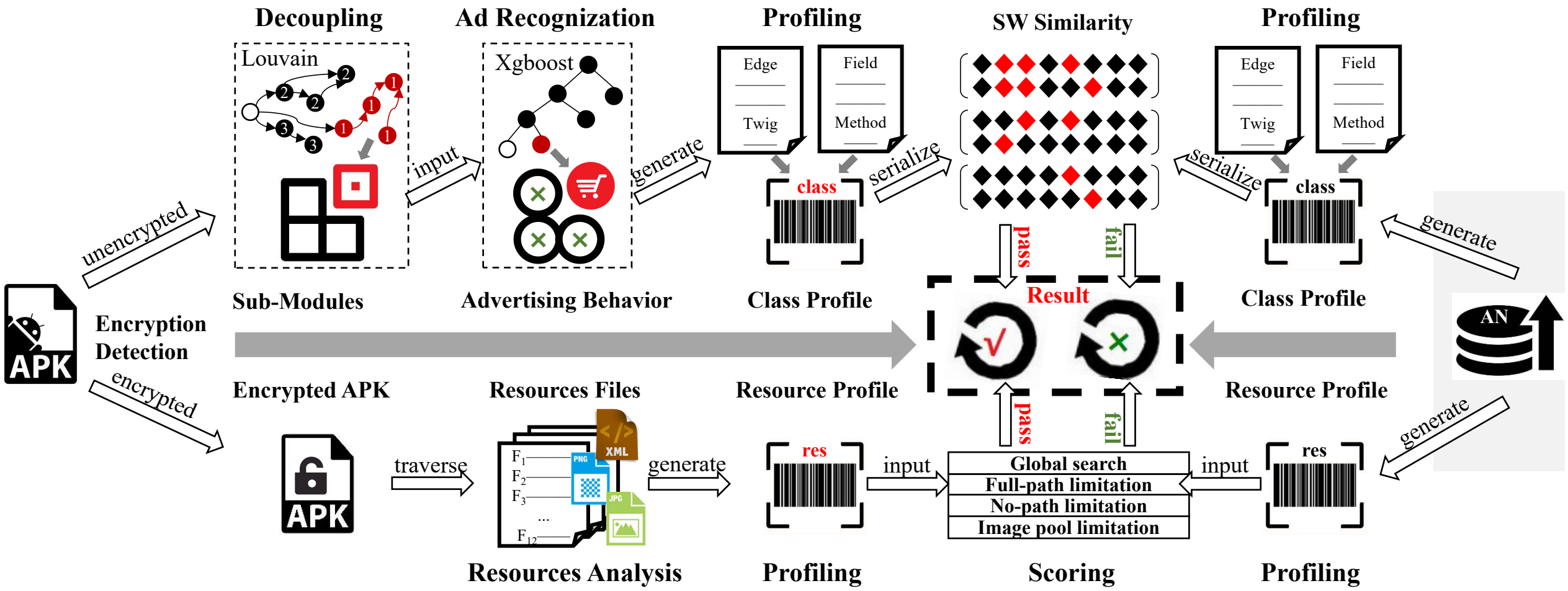
Table 5: The key, universality, differentiation and insensitivity of each feature.

Feature	Key	U_k	D_k	IS_k
F_1	AndroidManifest.xml_manifest_package	1.0000	0.7294	0.7913
F_2	AndroidManifest.xml_manifest uses-permission_android:name	0.5385	0.1122	0.3215
F_3	AndroidManifest.xml_manifest application provider_android:authorities	0.1186	0.0517	0.7953
F_4	AndroidManifest.xml_manifest application receiver intent-filter action_android:name	0.1602	0.0648	0.7222
F_5	AndroidManifest.xml_manifest application activity_android:name	0.3960	0.4886	0.4251
F_6	AndroidManifest.xml_manifest application service_android:name	0.1609	0.3541	0.5326
F_7	AndroidManifest.xml_manifest application receiver_android:name	0.1943	0.3261	0.6667
F_8	res\values\values.xml_resources dimen_name	0.1350	0.1666	0.1018
F_9	res\values\values.xml_resources string_name	0.3770	0.2886	0.2112
F_{10}	res\values\values.xml_resources declare-styleable attr_name	0.1302	0.3541	0.1628
F_{11}	res\values\values.xml_resources style_name	0.2529	0.1062	0.3273

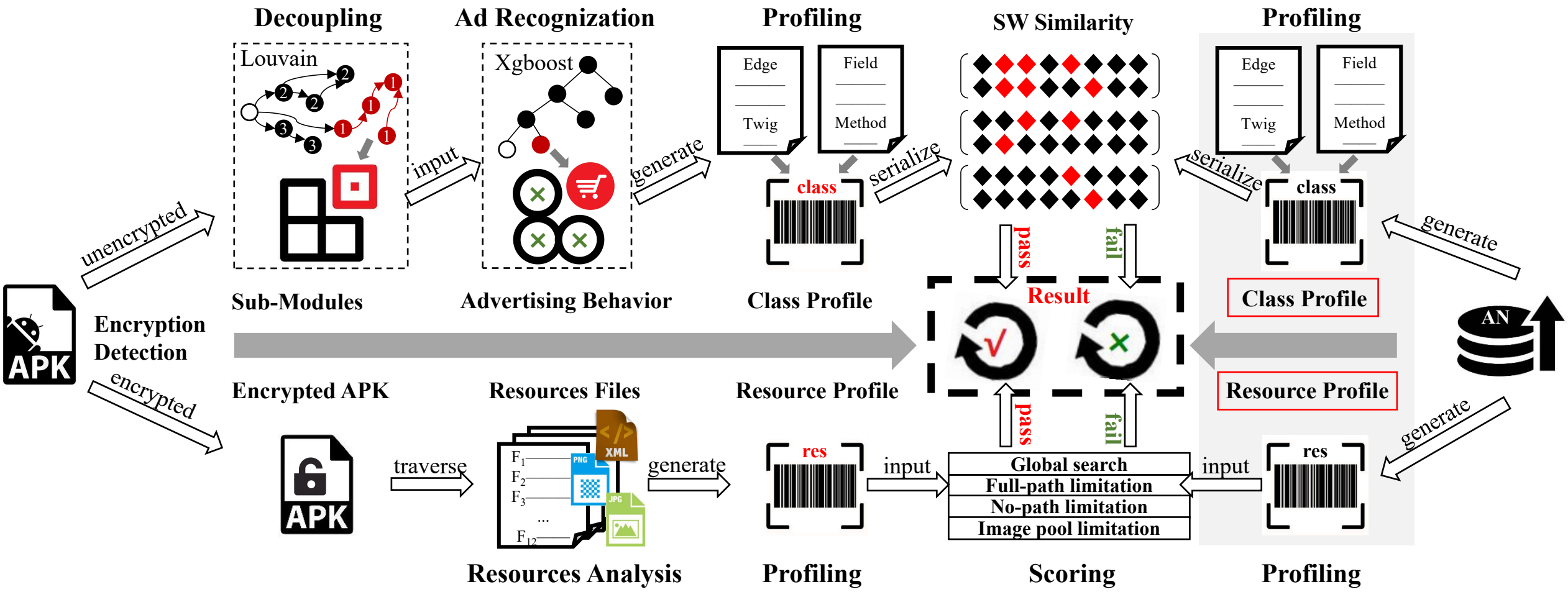
Global search | **Full-path match** | **No-path match** | **Image pool**

Comprehensive scoring

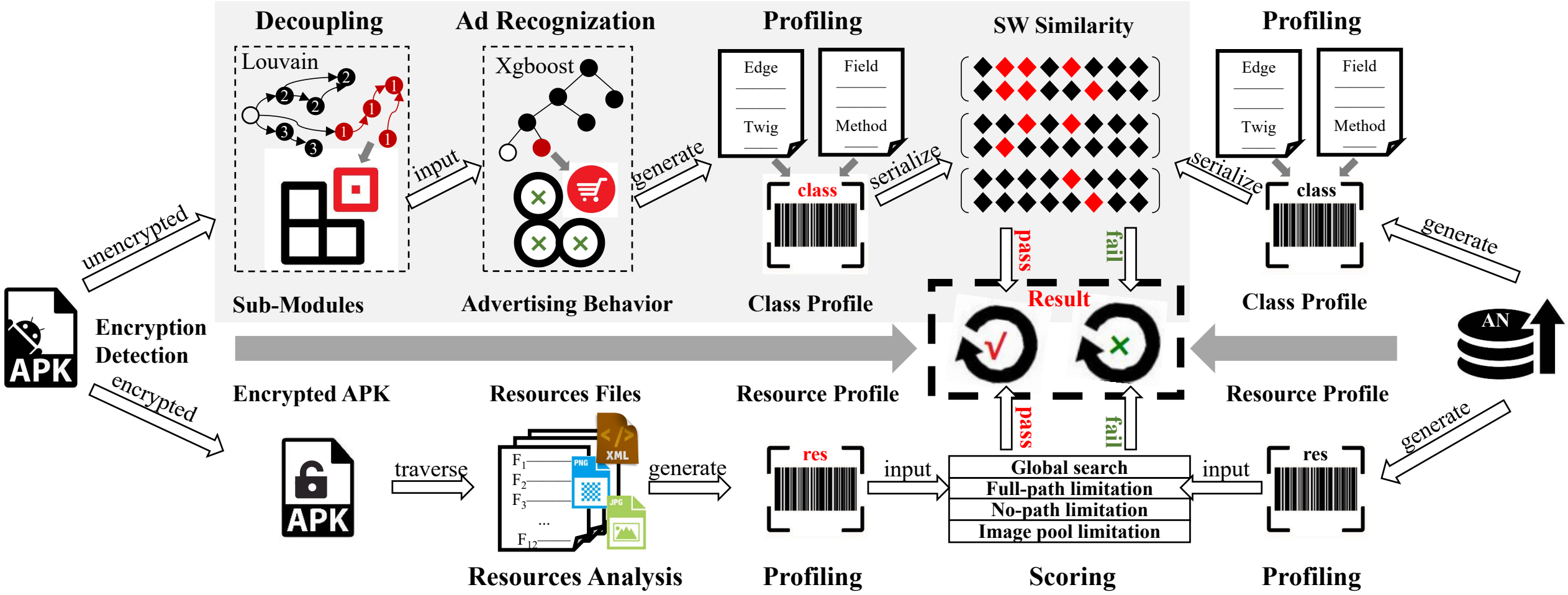
Architecture of ANDetect



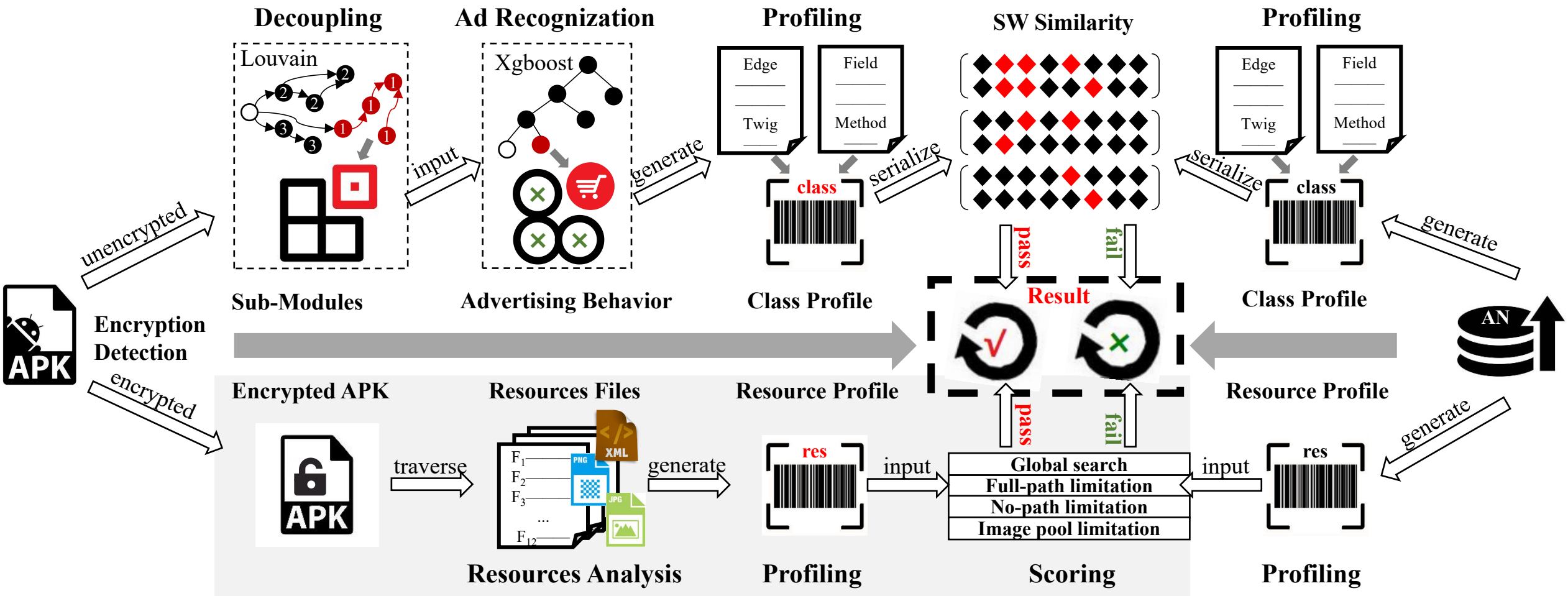
Architecture of ANDetect



Architecture of ANDetect



Architecture of ANDetect



Datasets

- **Evaluation metrics**
- **Fine-grained True Positive(TP-FG)**
- **Coarse-grained True Positive(TP-CG)**
- **Precision**
- **Recall**

- **Datasets**
- **Android Application Dataset(20,000 adwares & 125,401 unlabeled APPs)**
  - Low-confidence labeled datasets
  - High-confidence labeled datasets
- **Ad Networks Dataset(833 Ad libs & 2758 non-ad libs)**

Evaluation

Performance of ANDetect:

- Leveraging a 2:1 mix of non-encrypted and encrypted applications.
- Compare with **LibD**, **LibRadar** and **LibScout**.
- **Low Precision:** extra ad libs
- **High Recall:** competitors' misclassification of *com.google.android.gms.ads*

Table 1: The performance of ANDetect compared with other three tools in the experimental environment.

Tools	TP-CG	TP-FG	Precision	Recall	Time(s)/AP
LibD	46.08%	78.31%	99.72%	78.31%	58.14
LibRadar	48.04%	77.86%	99.50%	77.86%	29.37
LibScout	25.49%	72.64%	99.54%	72.64%	74.54
ANDetect	95.10%	95.97%	95.84%	95.97%	34.07

Table 2: The performance of ANDetect compared with other three tools in the real-world environment.

Tools	TP-CG	TP-FG	Precision	Recall	Time(s)/AP
LibD	47.40%	62.67%	98.73%	62.67%	153.45
LibRadar	43.35%	59.69%	98.26%	59.69%	65.11
LibScout	32.37%	53.47%	99.79%	53.47%	246.51
ANDetect	93.64%	93.08%	97.29%	93.08%	69.24

Evaluation

Robustness in encrypted APP detection:

- Randomly select 100 non-encrypted APPs with **resource confusion**.
- Tagged by the combined outputs of LibScout, LibD and LibRadar (**Low-confidence**)
- Identify the ad libs through **ANDetect's encrypted APP detection module**

- **Resource confusion properties:**
  - *res* directory is absent;
  - over half of the file names in *res* comprise less than 5 characters;
  - over half of the file names in *color* or *xml* comprise less than 5 characters;

Table 3: The result of encrypted application detection in resource confused applications.

Dataset	size	TP-CG	TP-FG	Precision	Recall
AP'_{adware}	100	94.00%	86.24%	81.84%	86.24%
AP'_{truth}	100	94.74%	91.50%	72.27%	91.50%

Evaluation

Novel ad libraries:

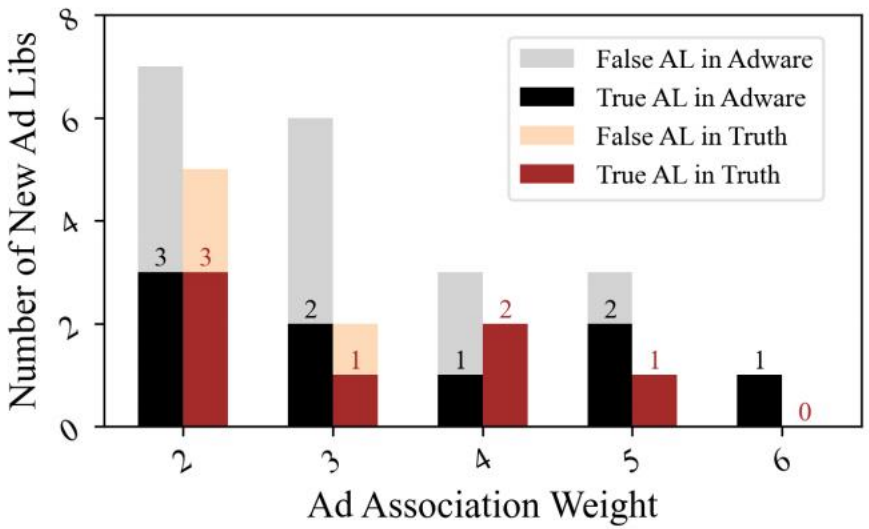


Figure 5: The novel ad libraries detected by ANDetect. The number of genuine ALs and false positives from AP_{adware} and AP_{truth} are given respectively.

- ANDetect’s non-encrypted application detection module
- Discern the true ad libs:
  - ◆ define ad association weight
  - ◆ search new lib name in Google
  - ◆ compute associate weight
- Allow for false positive to identify more novel ad libs
- **16** novel ad libs

Evaluation

Novel adwares:

- Label 20,000 adwares using VT
- Relationship between ad libs and adware

Table 4: The candidates of malicious ALs.

Candidate ALs	VirusTotal's Label	Relevance
com.adsmogo	Adware.ADWARE/ANDR.AdMogo	0.50
net.youmi.android	Android.Adware.Youmi.A	0.47
com.mopub.mobileads	AdLibrary:MoPub	0.45
com.inmobi.ads	AdWare.AndroidOS.Inmobi	0.80
com.nd.dianjin	Android.Dianjin.A (PUP)	0.50
cn.smartmad	Android.Adw.SmartMad	0.12
cn.domob.android	Android.Domob.A (AdWare)	0.26
com.kuguo.ad	Andr.Adware.Kuguo-4	1.00
com.jirbo.adcolony	Adware/AdColony!Android	0.50
com.sixth.adwoad.mraid	ADWARE/ANDR.AdsWo.FAN.Gen	0.43
org.iqiyi.video	AdWare.AndroidOS.IqiAd.a	1.00
com.revmob	Adware.Revmob.1.origin	0.20

Table 7: The brand-new adware detected by ANDetect. The sha256 of each adware and malicious ALs it contains are provided.

sha256	Malicious ALs
0002CA163461C26D3EA1D7384DB1833D4002CA679D0BA1906494E1BB24A95896	com.mopub.mobileads
00120DAD331324ED394118079BCB91D2B3F69C42E3C5135303DA50272FAE45F7	com.inmobi.ads;com.mopub.mobileads
00154456B10D6BF1F16DA14267D6063DAABEC20FFD28ABAB942D7BBD266D430F	com.mopub.mobileads
0134CE0FC74B7DDD800CF772A1742F7966FA4D5762359D2E6D15BED8728CC779	com.mopub.mobileads
0120397B85BFF474D554F82BEC39A5AB9827F93B68FD505AF89B935497860BAE	com.mopub.mobileads
0137F1EBE1F012BF454F78FE2452C1D39E0105AD0A00AC9E90F799EC96D09406	com.mopub.mobileads
7C90D85F5E08639CA32A5473CBDBE1B60334634B5CDB5AA7274FFF7A3B7B9BEA	com.mopub.mobileads
7C98A973971958D54D69D04C504A14EE8BF9CD709882BA544743F6D272C18955	com.mopub.mobileads
7CC341B217FA38C3C12BFA008166F6F7229ABE3CB765D4AD5455925D5EB20632	com.mopub.mobileads
7CE87F293FDDCA68C5786976B7F9AF9CF37D2598B383424086066BF41814372F	com.mopub.mobileads
7CF7F9D84FEA8DE117088FDAA4A3148DE301AFB21EA64BF3DEF244F93FF1B637	com.mopub.mobileads
7CD48EA847234C09C2D5EB60C8950005AE38674DDE24D9F5EBBD6043D533A10D	com.mopub.mobileads
7CC61FA9A5E041957CFA9F5434E540624988D6E139DCC71D85348F793AFEA475	com.mopub.mobileads
A4E9D3649543CF6246EEA45BFE15894C2C0430FDB4CA2787B3DEF2CECDF84521	com.mopub.mobileads
A4FD6B5E2E6F0FE0CBE89230FA9122647FF22D2DB814C247B3F8532614E347B0	com.mopub.mobileads
A552F35988EE74751A650C5AEB74FB1B2EEFD1231851D19AD23DB91923BBC7A8	com.mopub.mobileads
A56A45FBA78F4E785593F747E6FE9A7F3BFF7A94EAF1D08428D67959629C538F	com.inmobi.ads;com.mopub.mobileads
A58772EAAA85CB507003199F0129AB918BCEB6120CDD202C6CE4EBAC52A6C5BE	com.mopub.mobileads
A5A46FDC7879DC0210ECEF7ECC29D2369A1B00165BBAA668115B6BD863AF600C	com.mopub.mobileads
A5CAFB80BCF17912772C72C3DAB41D439B8EA06DD42A174A01C6864491F1F22D	com.mopub.mobileads
A5C2D903596DA9A8880BA97B3B9A9783391B1552D3B605FFF9134C3ED4B82E0E	com.mopub.mobileads
B33A5CD6AF62E8865DD7C2127CDDB9B96813AE6DE365EC6F0DF135E127C35BE6	com.inmobi.ads
B33949FA901C3227FD46416201AB2E219AEE15C804A8471D8213551E44E1C33A	com.mopub.mobileads
B34E6C6178B1FC6A3EAA4BFF12EB9BB6493F7F44E39BAA8ED44C91890BC92B04	com.mopub.mobileads
B3548B219682E7E739E32DE22B35BA1997EF3B6C74D1E7C6218F907059A29771	com.mopub.mobileads
B3C2FD36A1CBC3E921DAC8072E3E907A4E82E25FF223133963A3872A15A9B4DB	com.mopub.mobileads;com.inmobi.ads
B40C[illegible]46B73C[illegible]4421[illegible]367[illegible]50A[illegible]80D6[illegible]	com.mopub.mobileads
B3E8[illegible]A[illegible]5182[illegible]4[illegible]81[illegible]05[illegible]A09[illegible]A[illegible]5982[illegible]	com.mopub.mobileads
B3EE48851086FE4F66186AF69461FD8A559D4AC87E8596AA2DF06F877F28005A	com.mopub.mobileads
B40C0F97811AE225EAE17BC8DE6ED3188A2F7219652A495ED161C23F35FEE55A	com.mopub.mobileads
B4215B5FF02FDC76788062ED7C1CAA69A2C7CCD5B4031E9F51DAC5CB7DEFFEC2	com.mopub.mobileads
B42CD94A973F1A2CC48B324EFF9F7E7648EA01C9B6FA11489EBBC6BA2754A855	com.mopub.mobileads
7D042F1D52DD9BF078B04353340F63753A14A62D5FFE838AA4B36AF7EAB27E9D	com.inmobi.ads;com.mopub.mobileads
7CE4CB61A91D3A960C14D996E24729965ACCE2BD465B9941C2479186ADF2D170	com.mopub.mobileads;com.inmobi.ads
7CF0724873BC4464FCF0D8F99D675DCB22F1CBA97017BD8545E44FDDAAE0CD52	com.mopub.mobileads
7D39817880B7ACAAFA7A05B6864D30CC89CD4DAC16AE64235FDD15DAA82178DE	com.inmobi.ads
7D1247A2391A9A9FB08C1F77682ED16AE8885C1D10385A1D7D7B788AF0A2A2A5	com.inmobi.ads
10856950A97C189B5458B537BE2DB5555202A50B15345E97F1D733DA97EE21DF	com.mopub.mobileads
A5EBC5A1052546C781513698F8BF1E6ADF2F87B0AD203C8282C07F93C4F68F1D	com.mopub.mobileads
A5DCBD35B67D47BCA804225A34DCFF3E2AFBD7544BBEFDAB54BBCD64D19A4125	com.mopub.mobileads
A5F5AB473F93203D78F078CAC4580561609FB40D35A39CB5B78DEDCDFDBCE843	com.mopub.mobileads
A624B80E79BDA4535298DB9A6B0BDC6A265BDA521926079687E4C79BDD325CE0	com.mopub.mobileads
B4744416B37E89D94B865BACF3CF8D76CF23B0504F571C89A1E4AEA38A2F7C6D	com.mopub.mobileads
11F0B6F0FC2BE6F5826D53EA5685307201FEDAED7FEB789123D2C07E03D03B68	com.mopub.mobileads
B4BBC4CB843C4501E79DEA682425AED436C2168A1EAC057099E0A8A2A1357729	com.inmobi.ads;com.mopub.mobileads
11EC680264465AA2E07D0E46DE62E6CDB449208BE53FB281DF39FE5E944D43F8	com.mopub.mobileads
B527EF6798ECA6FBD4C1A43FBB7BACCAC1ACDCEB144F504C314E91B5A7589939	com.mopub.mobileads
B436E30ADFFC7C4FD75FF38FC91014A181978348ADB6B04F2B34389CF6DA5926	com.mopub.mobileads
B491A9266968E3687CE258D0163A7B458D2B0DB5207FD843CD5C71CFAC63BA4C	com.mopub.mobileads;com.inmobi.ads
B3FEE4445B1D5F1D7D45FA98E0DF32FBB74ABFE8572E413B0EEE1E5767FB4A4D	com.mopub.mobileads
7CE6103DFC87DC72EB0BDFAAAC872467E7A893C6007D7C378CED6D8F3584F462	com.mopub.mobileads
7C9C220C986580B07A9A2E2241349E75F8F91FB310CF2F87EB9A100896C4FF22	com.mopub.mobileads
B43E0E61180E5CD03590CFC172D4627F0E321D4A58ED538901ABCE7023968657	com.mopub.mobileads

53 brand new adware

Summary

- **Limitation**
- Resource features do not have a clear differentiation effect like class features, which is reflected in **different versions** of ad libraries.
- Not all advertising libraries have meaningful resource features.

- **Conclusion**
- We introduce **ANDetect**, a pioneering tool designed to efficiently identify **ad libraries** in Android applications, overcoming the limitations of existing methods especially when dealing with **encrypted apps**. The experiment of ANDetect to over 140,000 apps has resulted in the detection of 16 novel ad libraries and 53 new adwares.

https://sites.google.com/view/andetect

Thank you!
Q&A

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS