

NoMoAds: Effective and Efficient Cross-App Mobile Ad-Blocking

Anastasia Shuba
UC Irvine



Athina Markopoulou
UC Irvine

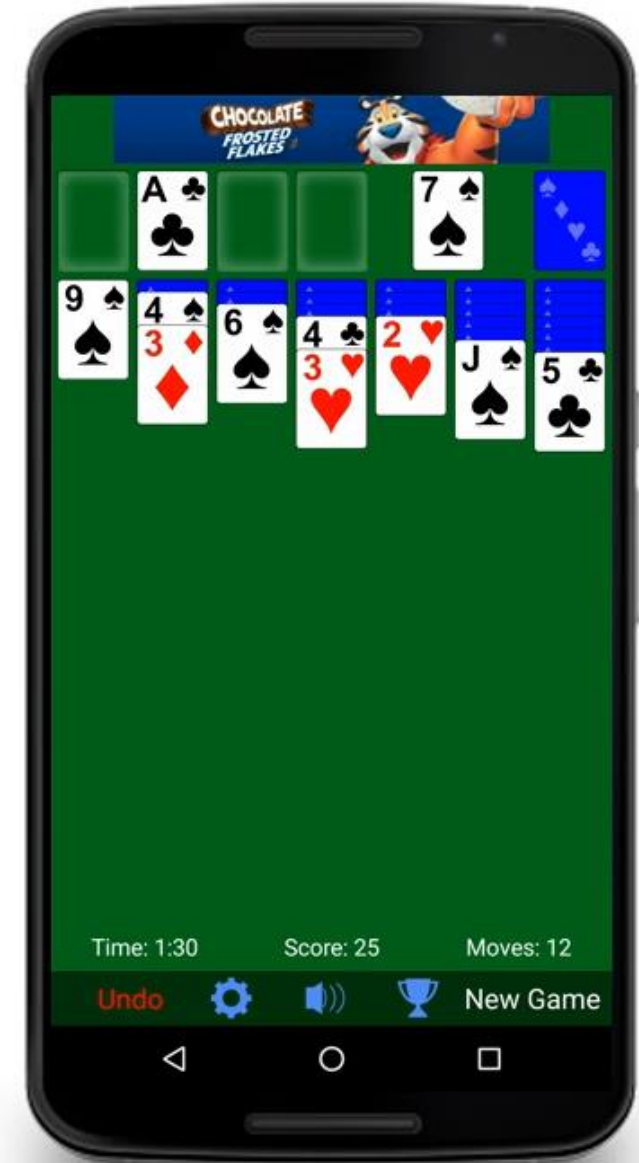
Zubair Shafiq
University of Iowa



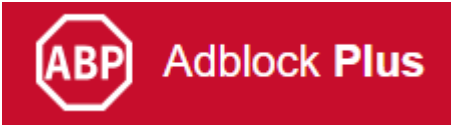
Motivation

Issues with Mobile Ads:

1. Intrusive
2. Overhead
3. Tracking
4. Malware



Background: Ads & Ad-Blocking



Get ad!



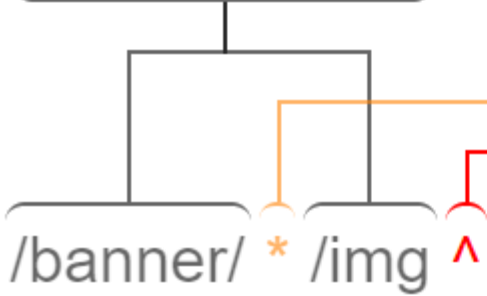
```
/adcode.  
/adcode/*  
/adcode_  
/adcodes/*  
/adcollector.  
/adcommon?  
/adcomp.  
/adcomponent/*  
/adconfig.  
/adconfig/*  
/adcontainer?  
/adcontent.$-object-subrequest  
/adcontent/*  
/adcontents_  
/adcontrol.  
/adcontrol/*  
openxtag.  
.org/ad-  
.org/ad.  
.org/ad/  
.org/ad_  
.org/adgallery1  
.org/ads-  
.org/ads/  
.org/ads_  
.org/adv/  
.org/exit.js  
.org/gads/  
.org/pops.js  
.ph/ads/  
.php/ad/  
.php/ads/  
.php?ad=  
.php?ad_  
.php?adsid=  
.php?adv=  
.php?adv_  
.php?affid=  
.php?clicktag=  
.php?nats=  
.php?zone_id=  
.php?zoneid=  
.p?adv=  
.pk/ads/  
.pl/ads/  
.popunder.js
```

EasyList (~64K rules)

Verbatim text
This text must be present in the address to be blocked.

Wildcard character
This stands for any number of characters.

Separator
The address must either end here or a separator character like ? or / has to follow.



Background: Related Work

	Detects Ads	Automated	Fine-Grained	Mobile-Specific	Cross-App	Runs on Device
DNS66, Disconnect, etc.	✓	✗	✗	✓	✓	✓
86% of Ad-Blocking Apps	✓	✗	✓	✗	✗	✓
AdblockPlus	✓	✗	✓	✗	✓	✓
[Razaghpanah et al. NDSS '18]	✓	✓	✗	✓	✓	✓
ReCon [Ren et al. MobiSys '16]	✗	✓	✓	✓	✓	✗
NoMoAds	✓	✓	✓	✓	✓	✓

NoMoAds Design

	Detects Ads	Automated	Fine-Grained	Mobile-Specific	Cross-App	Runs on Device
NoMoAds	Trained on mobile ads	Uses ML	Per-packet	Trained on mobile data	VPN-based solution	VPN app, no server

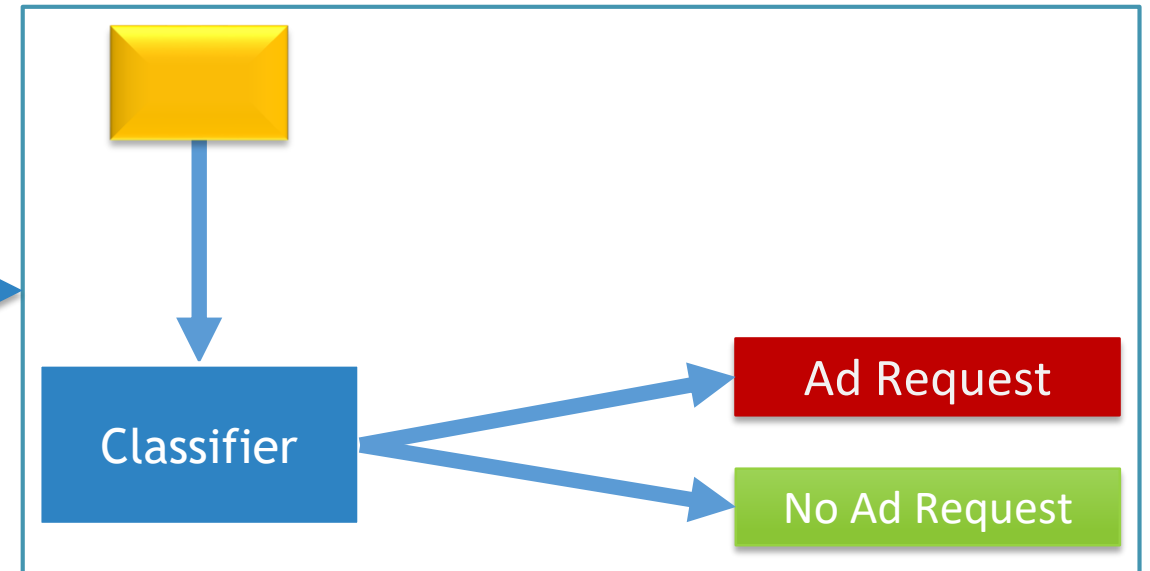
Effectiveness

Efficiency

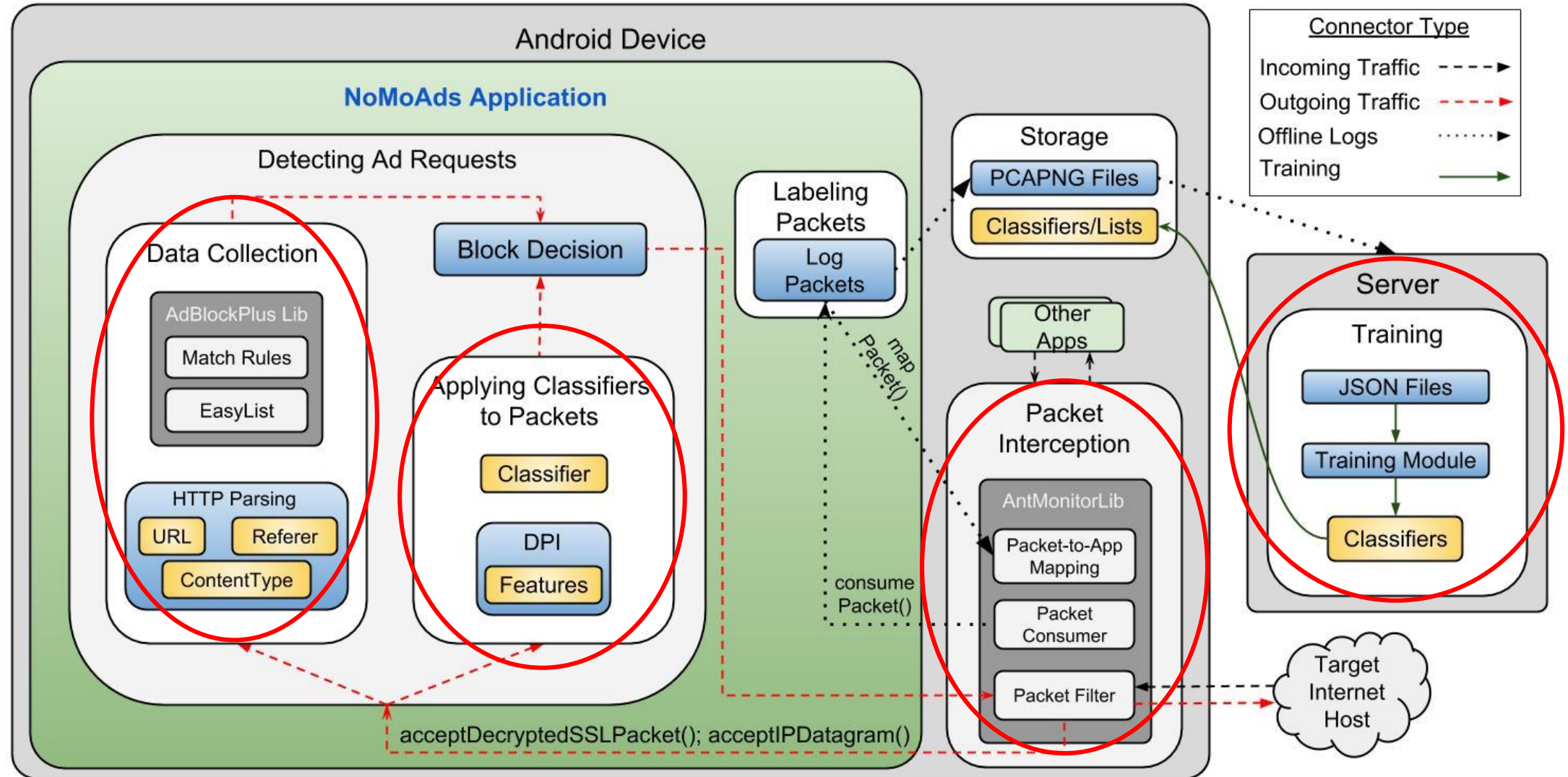
- ✓ Built on top of AntMonitor
- ✓ Avoid Java parsing



AntMonitor
[Shuba et al. '16]

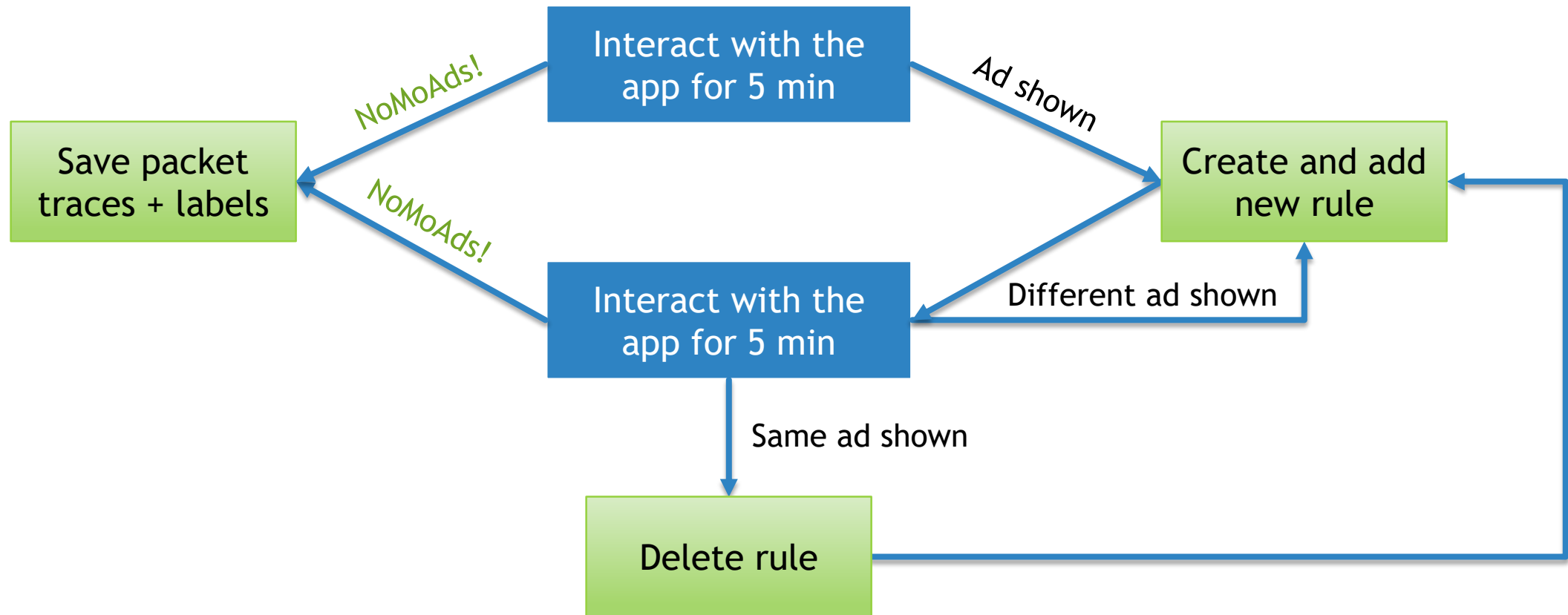


System Overview



Data Collection

- Tested 50 most popular apps that serve ads
- Used EasyList as a starting point

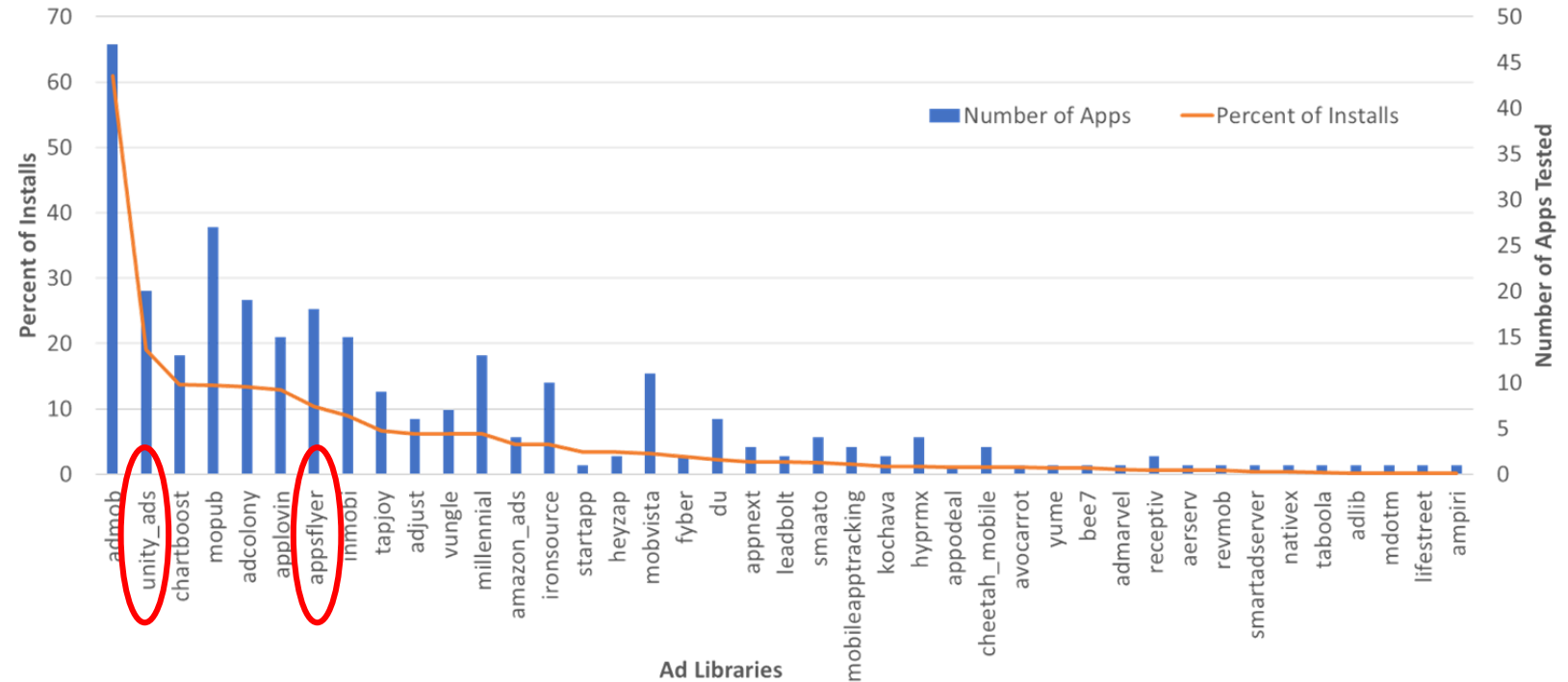


The NoMoAds Dataset

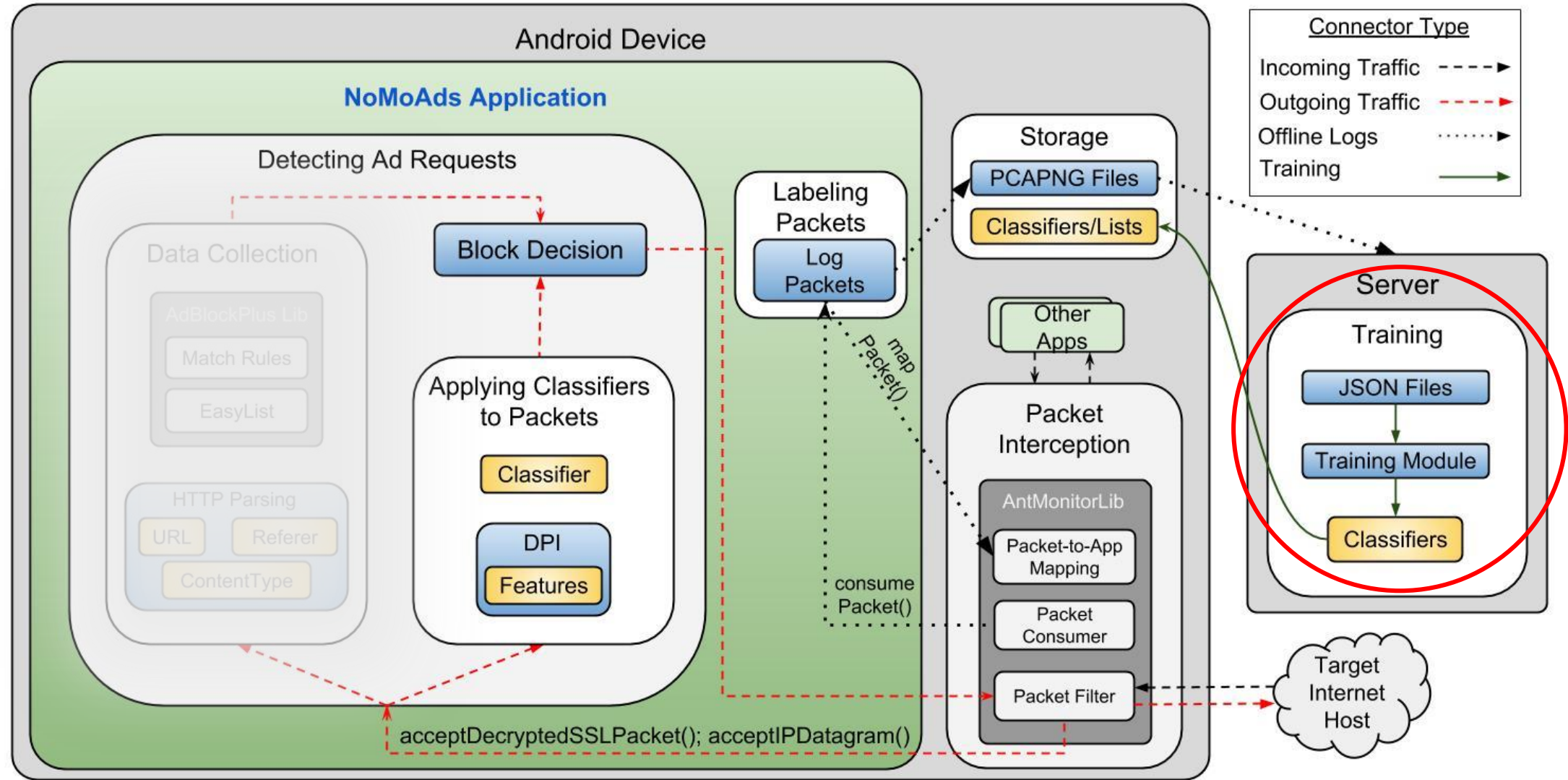
Available on our website!

<http://athinagroup.eng.uci.edu/projects/nomoads/>

- Apps: 50
- Packets: >15k
- Packets with ads: >4.5k
- Ad libraries: 41
- Good coverage: all libraries that account for at least 2% of installs
- First mobile ads dataset
- Using EasyList (and/or other lists) is not enough:
 - EasyList fails to detect >37% of ad requests
 - Detects ads generated by two of the most popular ad libraries (AdMob and MoPub)
 - Have desktop counterparts: Google owns AdMob, and Twitter owns MoPub
 - Fails to detect ads generated by libraries such as UnityAds and AppsFlyer



System Overview

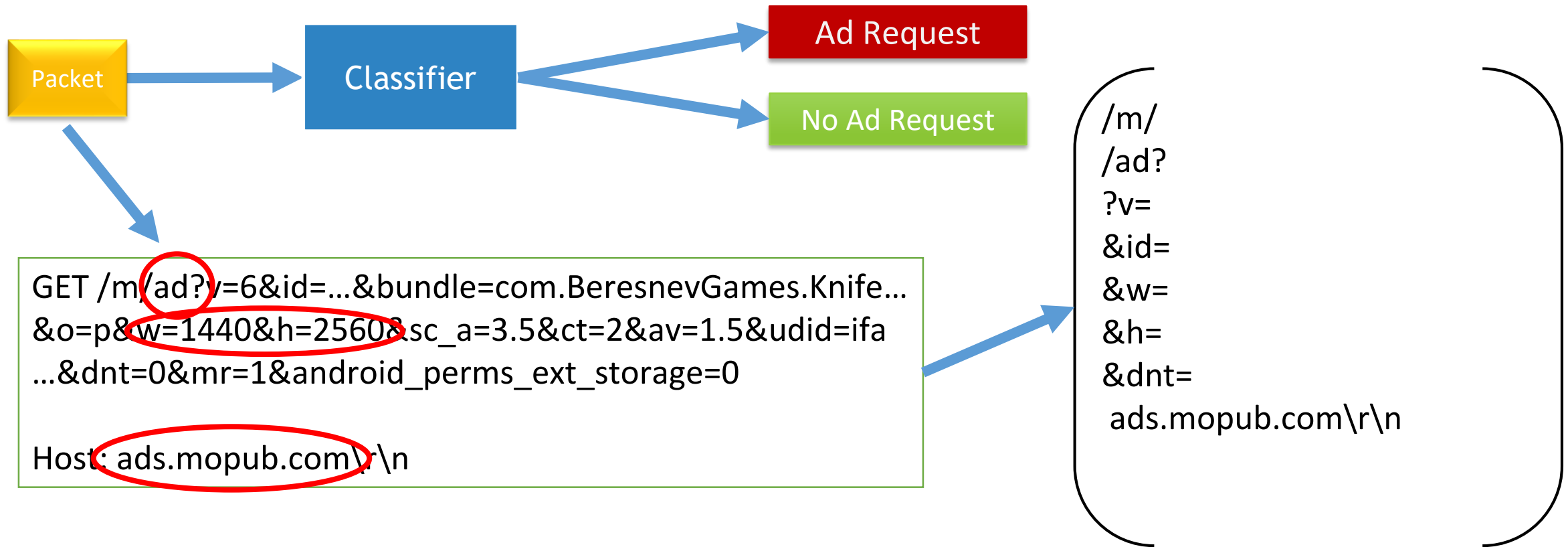


Training: Feature Selection

		Approaches Under Comparison	F1 score (%)	Number of Initial Features	Training Time (ms)	Tree Size
Ad-blocking lists		EasyList: URL + Content Type + HTTP Referer	77.1	63,977	N/A	N/A
		hpHosts: Host	61.7	47,557	N/A	N/A
		AdAwayHosts: Host	58.1	409	N/A	N/A
NoMoAds with Different Sets of Features		Destination IP + Port	87.6	2	298	304
		Domain	86.3	1	26	1
		...				

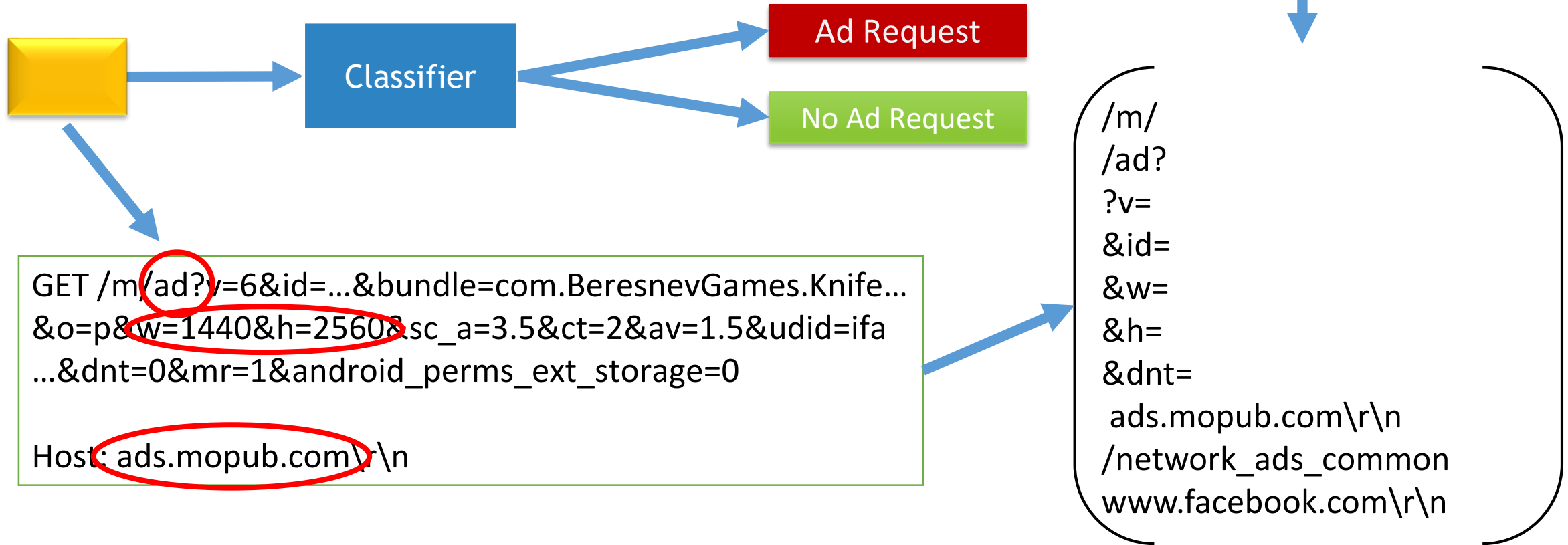
Training

- Built on ReCon - a system for detecting privacy leaks in mobile packets



Training

- Built on ReCon - a system for detecting privacy leaks in mobile packets



Training: Feature Selection

		Approaches Under Comparison	F1 score (%)	Number of Initial Features	Training Time (ms)	Tree Size
Ad-blocking lists		EasyList: URL + Content Type + HTTP Referer	77.1	63,977	N/A	N/A
		hpHosts: Host	61.7	47,557	N/A	N/A
		AdAwayHosts: Host	58.1	409	N/A	N/A
NoMoAds with Different Sets of Features		Destination IP + Port	87.6	2	298	304
		Domain	86.3	1	26	1
		...				

Training: Feature Selection

	Approaches Under Comparison	F1 score (%)	Number of Initial Features	Training Time (ms)	Tree Size
Ad-blocking lists	EasyList: URL + Content Type + HTTP Referer	77.1	63,977	N/A	N/A
	hpHosts: Host	61.7	47,557	N/A	N/A
	AdAwayHosts: Host	58.1	409	N/A	N/A
NoMoAds with Different Sets of Features	Destination IP + Port	87.6	2	298	304
	Domain	86.3	1	26	1
	URL	93.7	4,133	483,224	196
	URL+Headers	96.3	5,320	755,202	274

Training: Feature Selection

	Approaches Under Comparison	F1 score (%)	Number of Initial Features	Training Time (ms)	Tree Size
Ad-blocking lists	EasyList: URL + Content Type + HTTP Referer	77.1	63,977	N/A	N/A
	hpHosts: Host	61.7	47,557	N/A	N/A
	AdAwayHosts: Host	58.1	409	N/A	N/A
NoMoAds with Different Sets of Features	Destination IP + Port	87.6	2	298	304
	Domain	86.3	1	26	1
	URL	93.7	4,133	483,224	196
	URL+Headers	96.3	5,320	755,202	274
	URL+Headers+PII	96.9	5,326	770,015	277

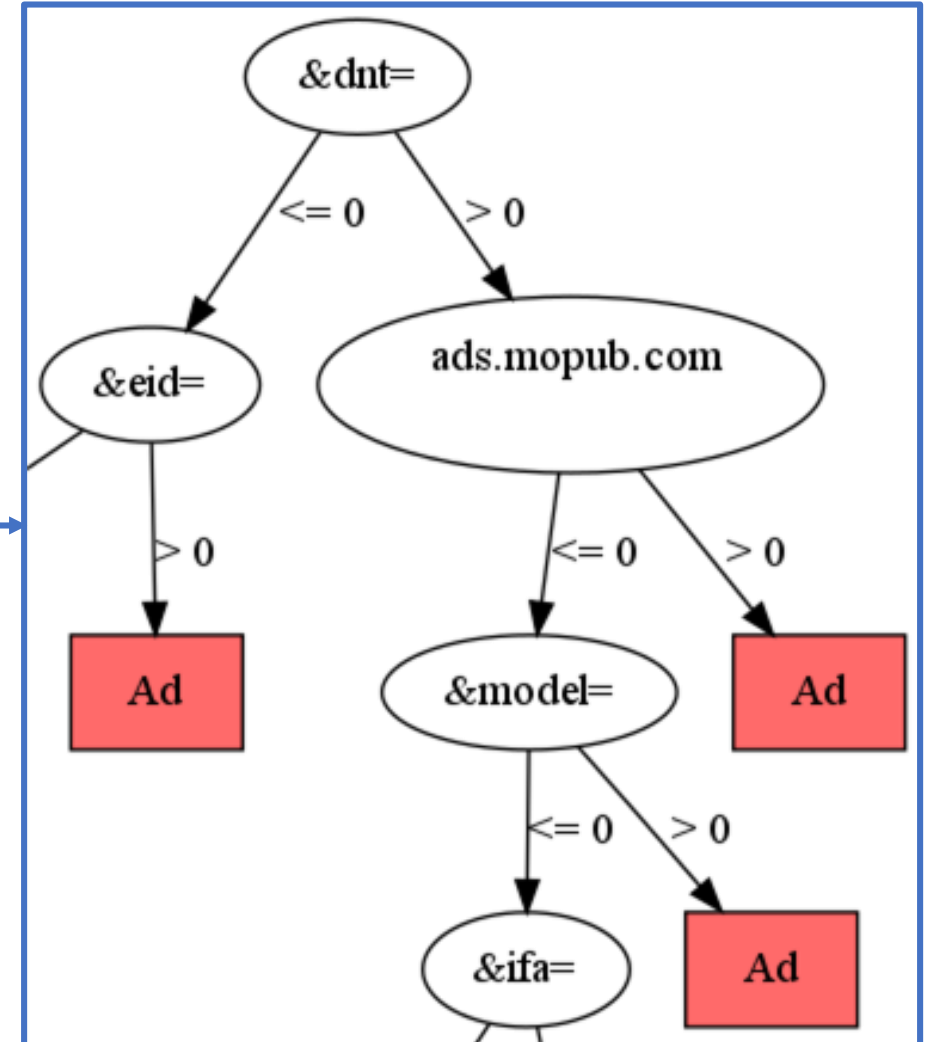
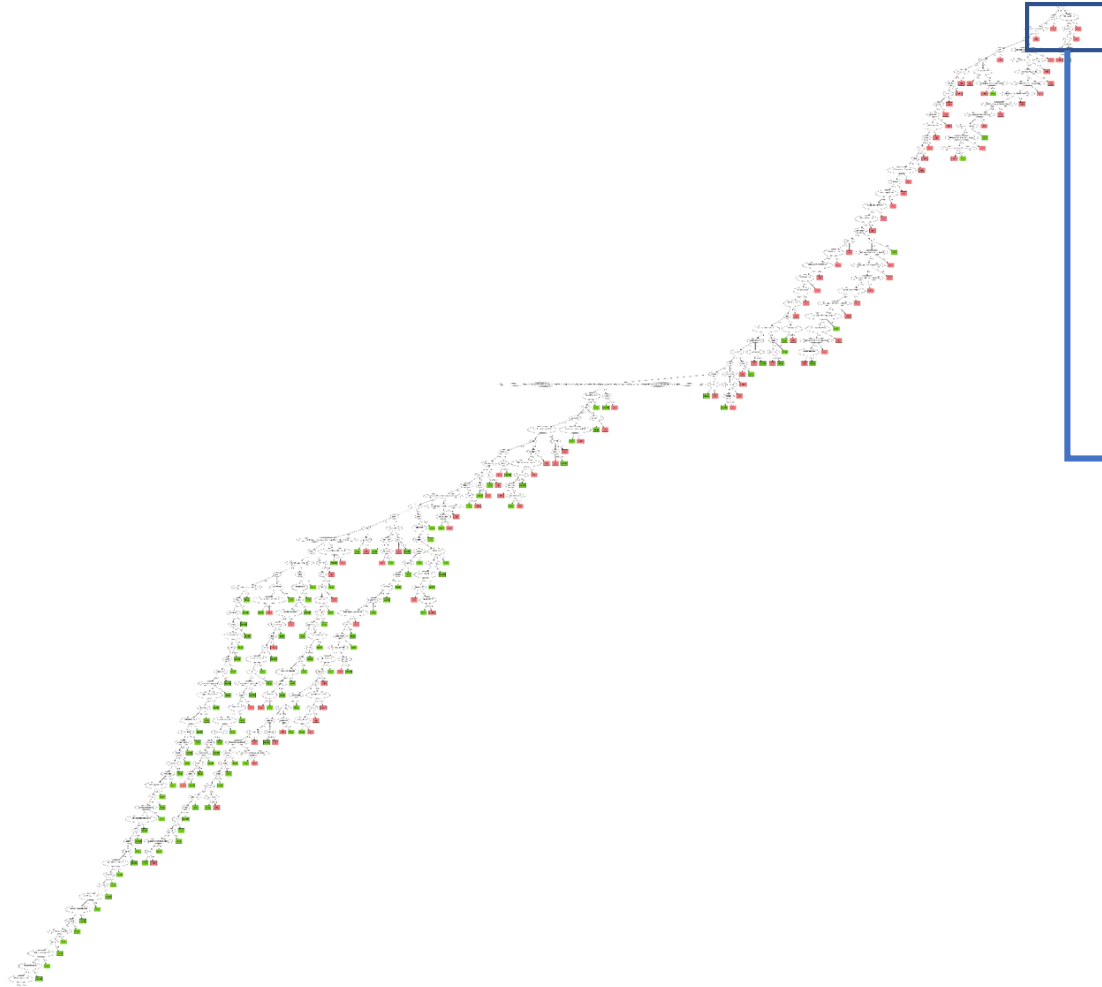
Training: Feature Selection

	Approaches Under Comparison	F1 score (%)	Number of Initial Features	Training Time (ms)	Tree Size
Ad-blocking lists	EasyList: URL + Content Type + HTTP Referer	77.1	63,977	N/A	N/A
	hpHosts: Host	61.7	47,557	N/A	N/A
	AdAwayHosts: Host	58.1	409	N/A	N/A
NoMoAds with Different Sets of Features	Destination IP + Port	87.6	2	298	304
	Domain	86.3	1	26	1
	URL	93.7	4,133	483,224	196
	URL+Headers	96.3	5,320	755,202	274
	URL+Headers+PII	96.9	5,326	770,015	277
	URL+Headers+Apps+PII	97.7	5,327	555,126	223

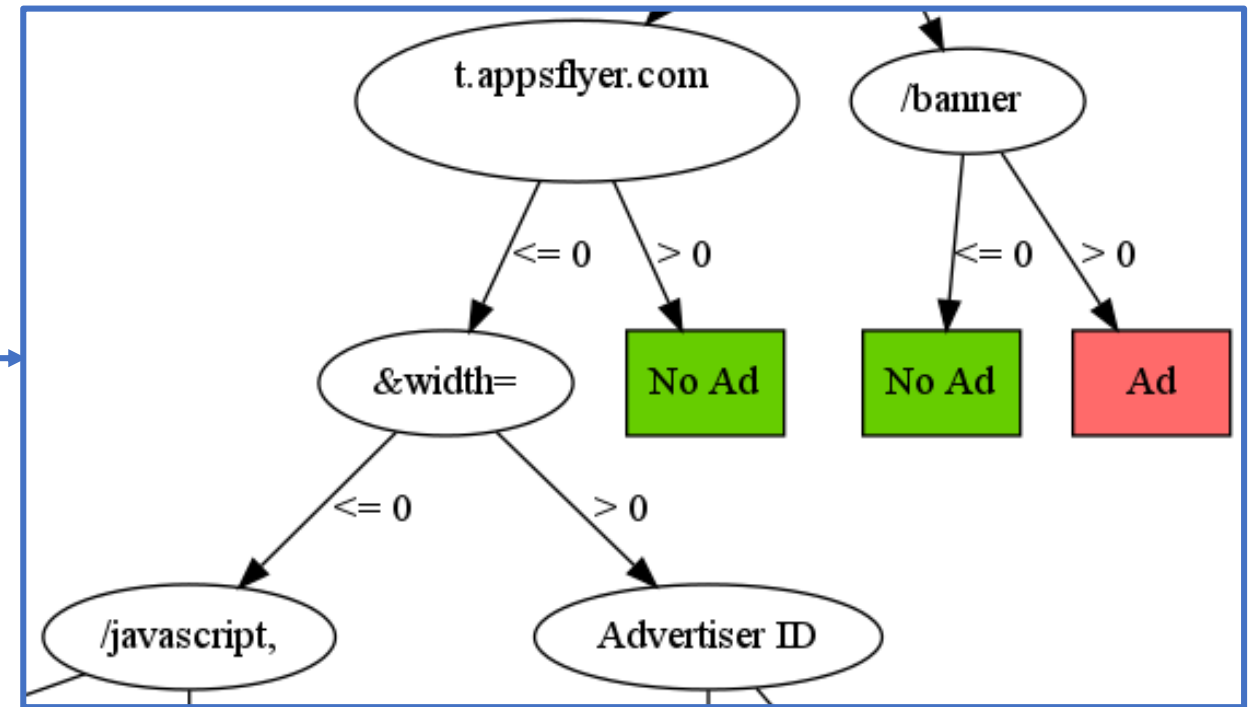
Training: Feature Selection

	Approaches Under Comparison	F1 score (%)	Number of Initial Features	Training Time (ms)	Tree Size
Ad-blocking lists	EasyList: URL + Content Type + HTTP Referer	77.1	63,977	N/A	N/A
	hpHosts: Host	61.7	47,557	N/A	N/A
	AdAwayHosts: Host	58.1	409	N/A	N/A
NoMoAds with Different Sets of Features	Destination IP + Port	87.6	2	298	304
	Domain	86.3	1	26	1
	URL	93.7	4,133	483,224	196
	URL+Headers	96.3	5,320	755,202	274
	URL+Headers+PII	96.9	5,326	770,015	277
	URL+Headers+Apps+PII	97.7	5,327	555,126	223
	URL+Headers+Apps	97.8	5,321	635,400	247

Training: The Decision Tree



Training: The Decision Tree



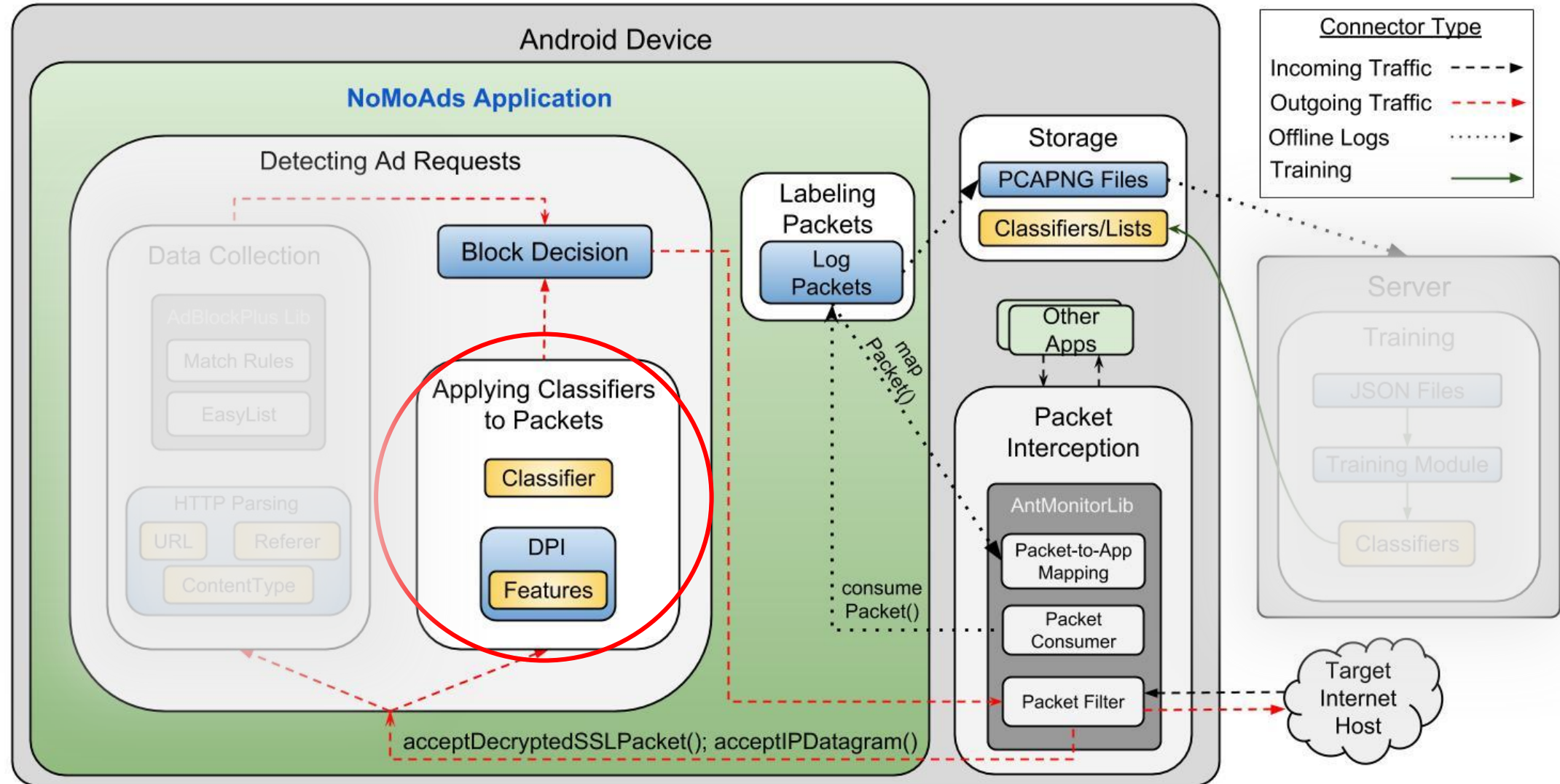
Training: Summary

- The more information we use, the better the F1 score
- Classifier trained on URL+Headers+PII perform well
- Split test and training sets:
 - Based on packets: 96.9% F1 score
 - Based on apps: 70% of apps have an F1 score \geq 80%
 - Based on ad libraries: 100% F1 score, even with library overlap below 100%

Code available on our website!

<http://athinagroup.eng.uci.edu/projects/nomoads/>

System Overview



Applying Classifiers to Packets

How to avoid this parsing step?

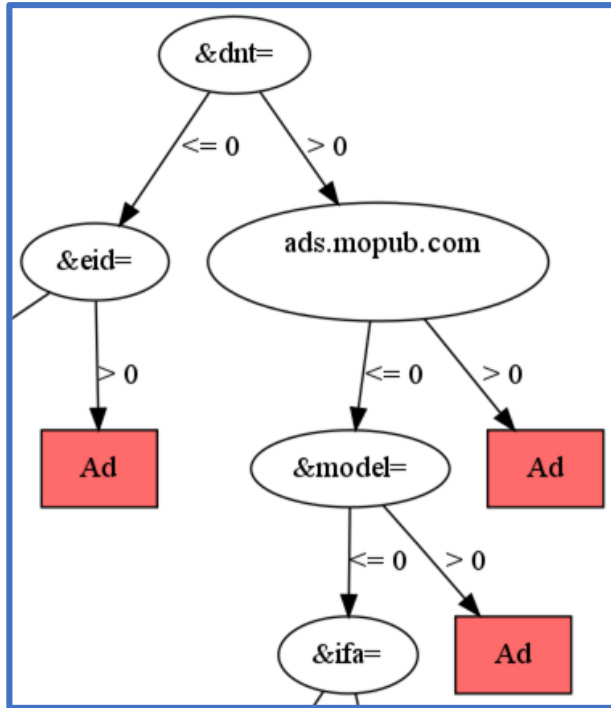
```
GET /m/ad?v=6&id=...&bundle=com.BeresnevGames.Knife...  
&o=p&w=1440&h=2560&sc_a=3.5&ct=2&av=1.5&udid=ifa  
...&dnt=0&mr=1&android_perms_ext_storage=0
```

```
Host: ads.mopub.com\r\n
```



```
GET  
/m/  
/ad?  
?v=  
=6&  
&id=  
&w=  
&h=  
&dnt=  
ads.mopub.com\r\n
```


Applying Classifiers to Packets



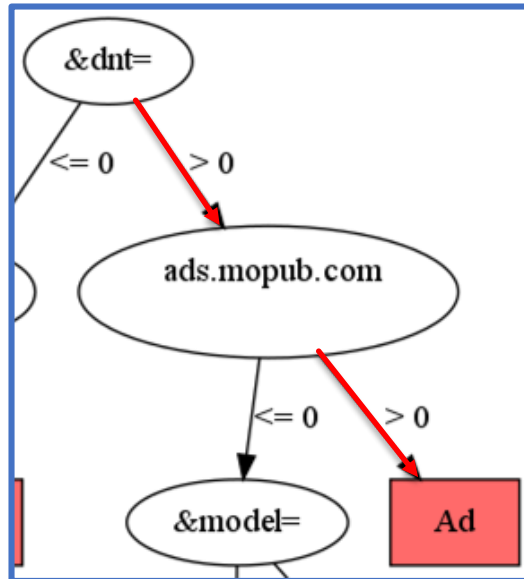
&dnt=
ads.mopub.com\r\n
&model=
&ifa=
/api/
/native/
cdn.outfit7.com\r\n
/soft/
okhttp/
&device_id=
com.smule.singandroid/
helpgrid1.ksmobile.com\r\n
boundary=

....

Applying Classifiers to Packets

GET /m/ad?v=6&id=...&bundle=com.BeresnevGames.Knife...
&o=p&w=1440&h=2560&sc_a=3.5&ct=2&av=1.5&udid=ifa
...&dnt=0&mr=1&android_perms_ext_storage=0

Host: ads.mopub.com\r\n



$\left[\begin{array}{l} \&dnt= \\ \text{ads.mopub.com}\backslash\text{r}\backslash\text{n} \end{array} \right]$

$\left[\begin{array}{l} \&dnt= \\ \text{ads.mopub.com}\backslash\text{r}\backslash\text{n} \\ \&model= \\ \&ifa= \\ /api/ \\ /native/ \\ \text{cdn.outfit7.com}\backslash\text{r}\backslash\text{n} \\ /soft/ \\ \text{okhttp/} \\ \&device_id= \\ \text{com.smule.singandroid/} \\ \text{helpgrid1.ksmobile.com}\backslash\text{r}\backslash\text{n} \\ \text{boundary=} \\ \dots \end{array} \right]$

Applying Classifiers to Packets: Evaluation

- First time on-device per-packet classification in real-time
- Setup
 - Timed how long prediction takes on a Nexus 6
 - Fed 10 HTTP packets of varying sizes (between 300-2000B), repeated 100 times
- Results:
 - Extracting features and applying the DT classifier: 2.96 ± 2.07 ms
 - Most of the delay from applying the classifier
 - Recent improvement: 1.87 ± 0.77 ms



Conclusion & Future Directions

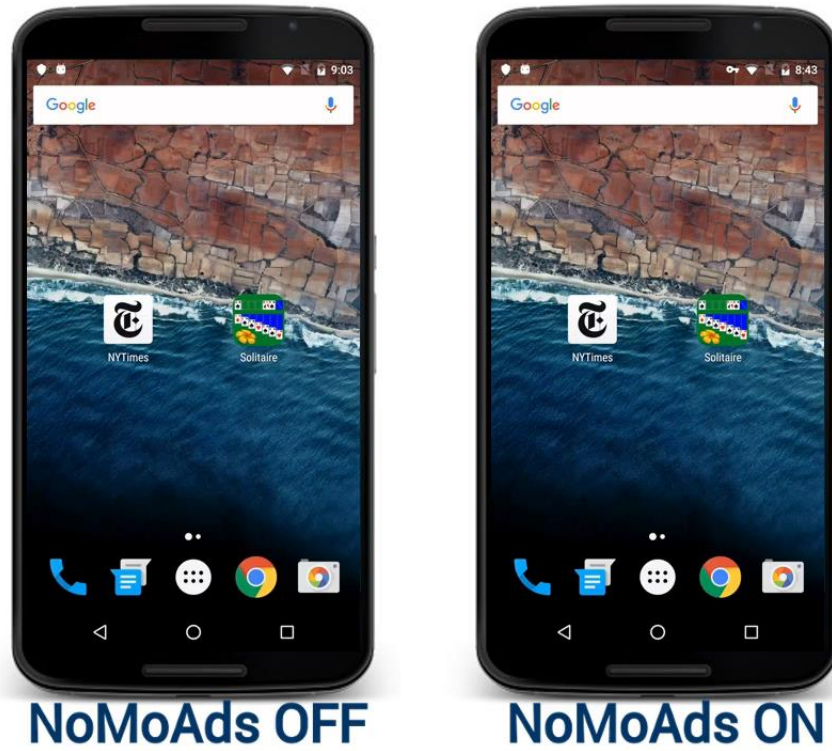
- NoMoAds: cross-app mobile ad-blocker
- Effective and efficient machine learning on the mobile device

Future Directions

- Larger dataset
- Detect app breakage
- Extend to trackers



Photo by Jimmy Nelson



Anastasia Shuba
ashuba@uci.edu

NoMoAds Code and Dataset Available at:
<http://athinagroup.eng.uci.edu/projects/nomoads>