StructSignature: Empirical Analysis of Android Binary File Layout

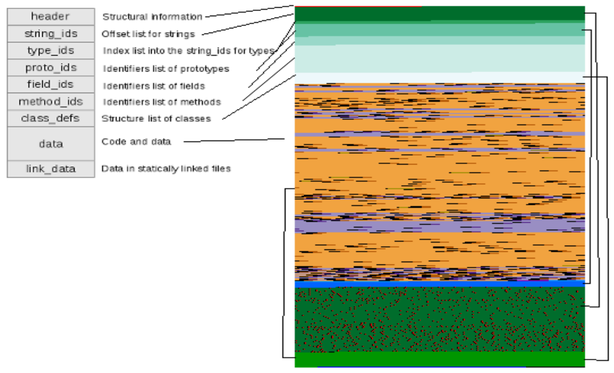
Hugo Gonzalez, Natalia Stakhanova, Ali Ghorbani

Canadian Institute for Cybersecurity (CIC), University of new Brunswick (UNB)

Motivation

- Different tools to create Android apps are available for Developers.
- .dex file contains traces of the tool used to produce it.
- Similar development environment produce .dex files with same structural layout.
- Individual developers or groups will use same tools consistently.
- WhatsApp dataset.

StructSignature: The tool



1. Analysis of .dex file

Element name	Representation	In Compact signature
'Header'	'H'	
'Link'	'L'	
'Strings_id'	's'	
'Type_id'	't'	
'Prototype_id'	'p'	
'Field_id'	'f'	
'Method_id'	'm'	
'Class_def'	'c'	
'Map'	'Z'	Y
'Strings in order'	'S'	Y
'Strings not in order'	'R'	Y
'Data'	'D'	Y
'Prototypes'	'P'	Y
'Interfaces'	'I'	Y
'Anno_dir_list'	'A'	Y
'Encoded_method'	'M'	Y
'Code_item_debug_info'	'd'	Y
'Encoded_value'	'V'	
'Class_data'	'C'	Y
'ExtraHeader'	'X'	Y
'Anno_set_item'	'E'	
'Anno_item'	'T'	
'BEGIN'	'>'	Y
'END'	'<'	Y

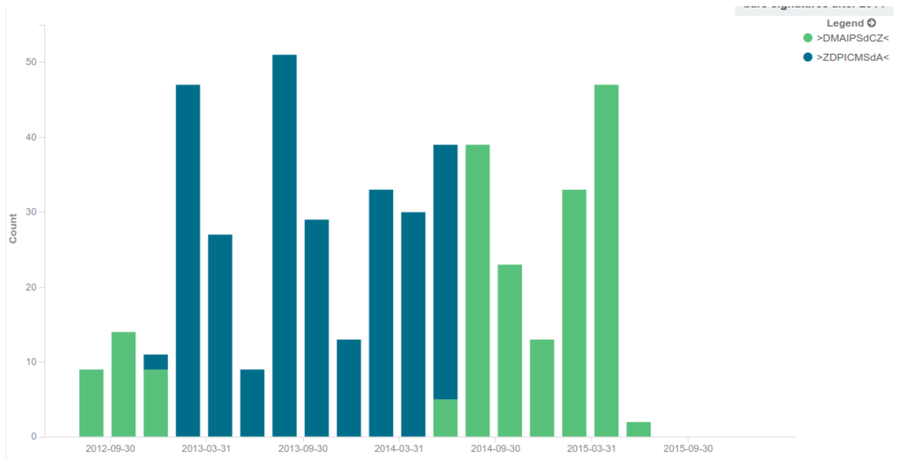
2. Structural Elements

3. Signatures

>ZDPICMSdA<
>DMAIPSdCZ<
>DMAIPSdCZ<
>DMAPIRCdZ<
>DMAPISCZ<
>DMAIPSCZ<
>DMAIPRCdZ<
>SDPIAdMCZ<
>SDPIAMCZ<

WhatsApp dataset

- 469 official releases where collected between 2012 and 2015.
- Two unique signatures consistent in time.
- Hypothesis:
  - New version of tools



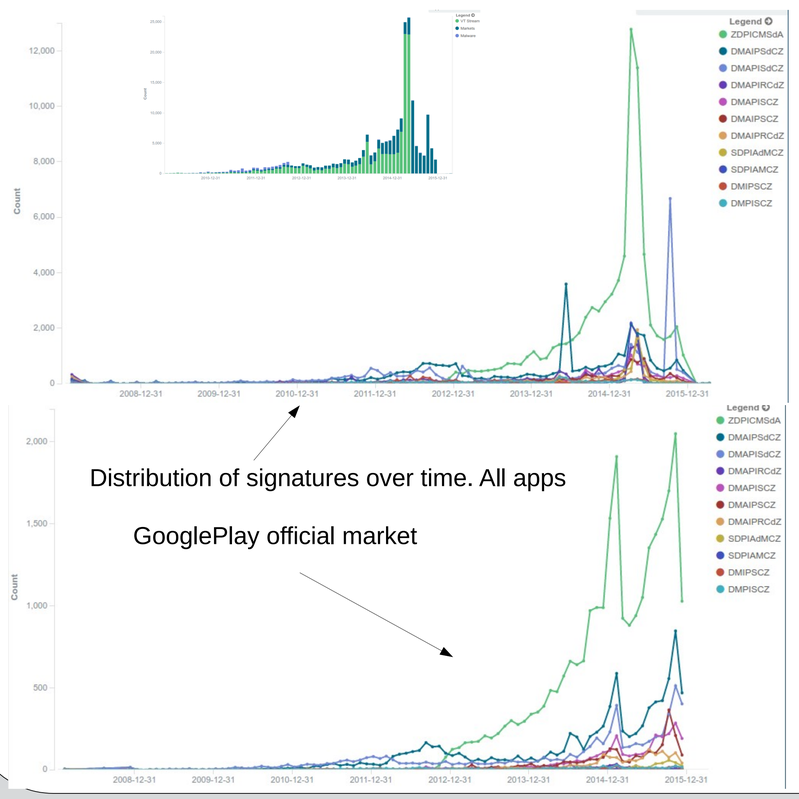
Dataset

200K

Table 2: Datasets used

Name	Apps	Description
Whatsapp	469	Whatsapp application. Different versions collected from official market since 2012 to middle 2015.
Official market	50,324	Apps crawled from official Googleplay market from late 2014 to early 2016.
Other markets	29,588	Apps from third-party markets Including: tencent (7, 145), slideme (6,343), xiaomi (5,953), aptoide (4,442), anzhi (1,749), fdroid (1,395), 360 (1,312) and appland (1,249).
Github Samples	46	Manually compiled samples from Github.
Malicious	6,334	This apps were flagged as malware, it includes Drebin dataset (5,560) and malware, adware and ransomware collected recently (778)
Suspicious	116,338	Apps collected from VirusTotal stream service. All these apps were submitted to te service to be analyzed
Total	203,099	

Results

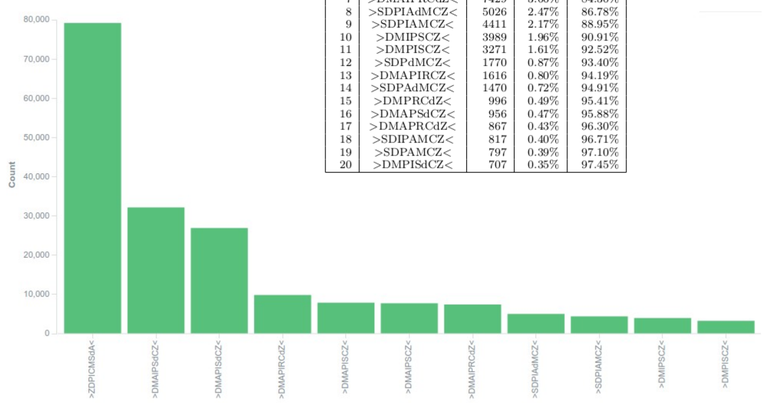


Distribution of signatures over time. All apps

GooglePlay official market

Table 3: Most common signatures

No.	Signature	Count	%	Cum. %
1	>ZDPICMSdA<	79187	38.99%	38.99%
2	>DMAIPSdCZ<	32178	15.84%	54.83%
3	>DMAPISdCZ<	26924	13.26%	68.09%
4	>DMAPIRCdZ<	9857	4.85%	72.94%
5	>DMAPISCZ<	7898	3.89%	76.83%
6	>DMAIPSCZ<	7745	3.81%	80.64%
7	>DMAIPRCdZ<	7429	3.66%	84.30%
8	>SDPIAdMCZ<	5026	2.47%	86.78%
9	>SDPIAMCZ<	4411	2.17%	88.95%
10	>DMIPSCZ<	3989	1.96%	90.91%
11	>DMPISCZ<	3271	1.61%	92.52%
12	>SDPdMCZ<	1770	0.87%	93.40%
13	>DMAPIRCZ<	1616	0.80%	94.19%
14	>SDPAdMCZ<	1470	0.72%	94.91%
15	>DMPRCdZ<	996	0.49%	95.41%
16	>DMAPSdCZ<	956	0.47%	95.88%
17	>DMAPRCdZ<	867	0.43%	96.30%
18	>SDPAMCZ<	817	0.40%	96.71%
19	>SDPAMCZ<	797	0.39%	97.10%
20	>DMPISdCZ<	707	0.35%	97.45%



- 10 signatures cover 90% of samples
- 20 signatures cover 97.4% of samples
- There are unique signatures on the dataset

Conclusions

- New feature based on .dex file layout.
- Not enough for malware detection but it can help to identify development tools and configurations.
- Can be used in a novelty detection system, unknown signatures are novelties or anomalies.
- As a feature in other classification or attribution systems.

Request StructSignature tool.