ISCX
Information Security
Centre *of* Excellence

Exploring the origin of Android Apps

Hugo González, Natalia Stakhanova, Ali A. Ghorbani

Faculty of Computer Science, University of New Brunswick

UNB

The Problem

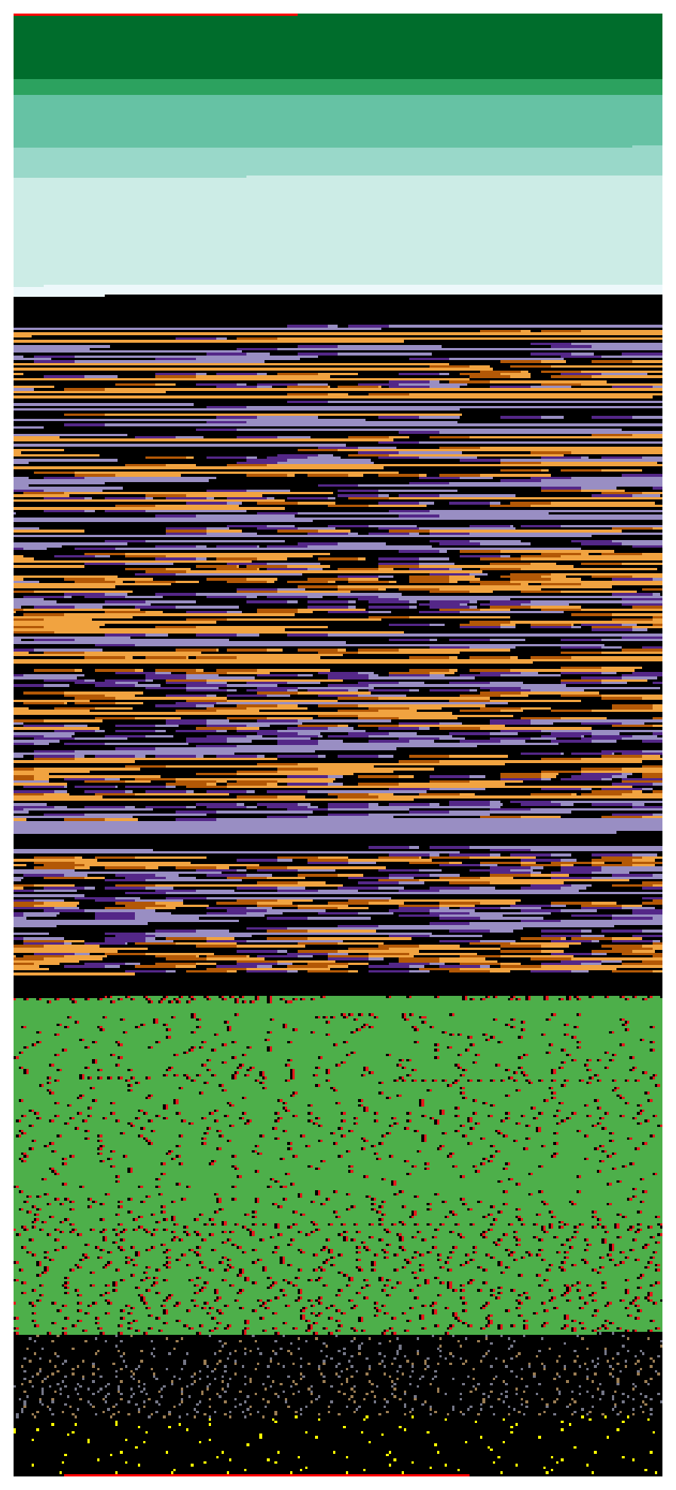
➢Hundreds of apps are sent to the markets each day, the majority are legit ones, for the rest we found malware and repackaged apps. Besides the maliciousness of the app, the repackaged ones affects the original authors because the revenue model or because the image of the creator.

Proposed Solution

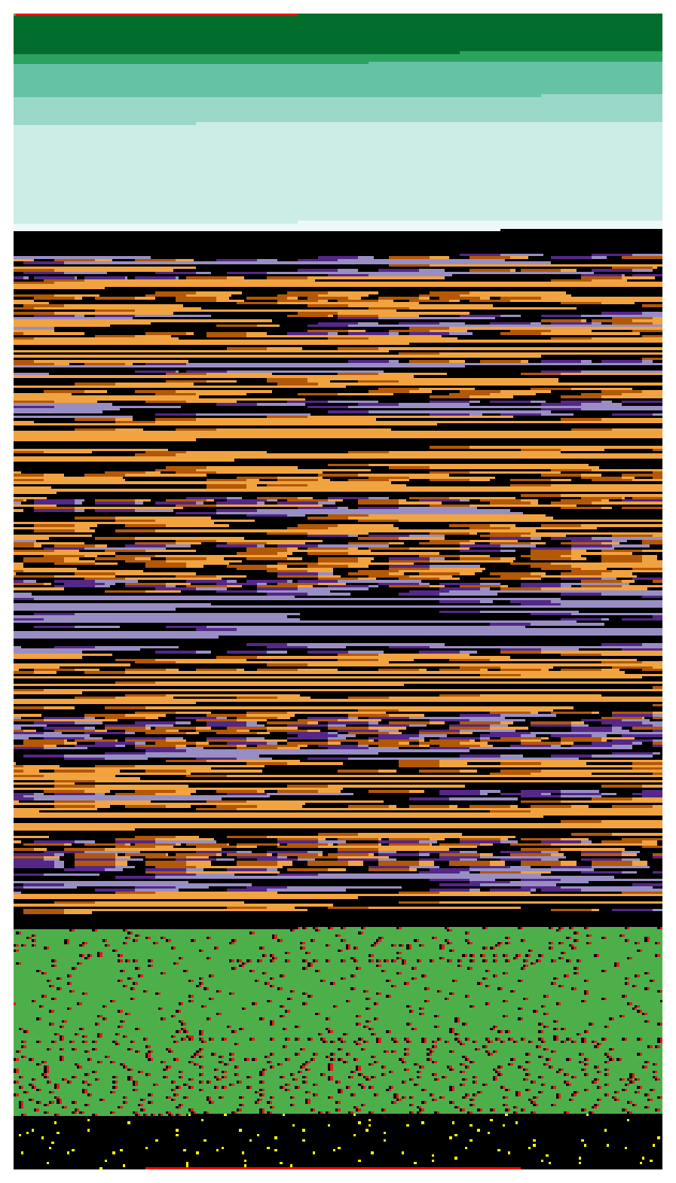
➢Analyzing the origin of an app answering the following questions, when was crated?, who signed it?, which tools were employed? We believe that by labelling the origin of an Android app we can flag it as suspicious, and related to the author.

.dex file structure

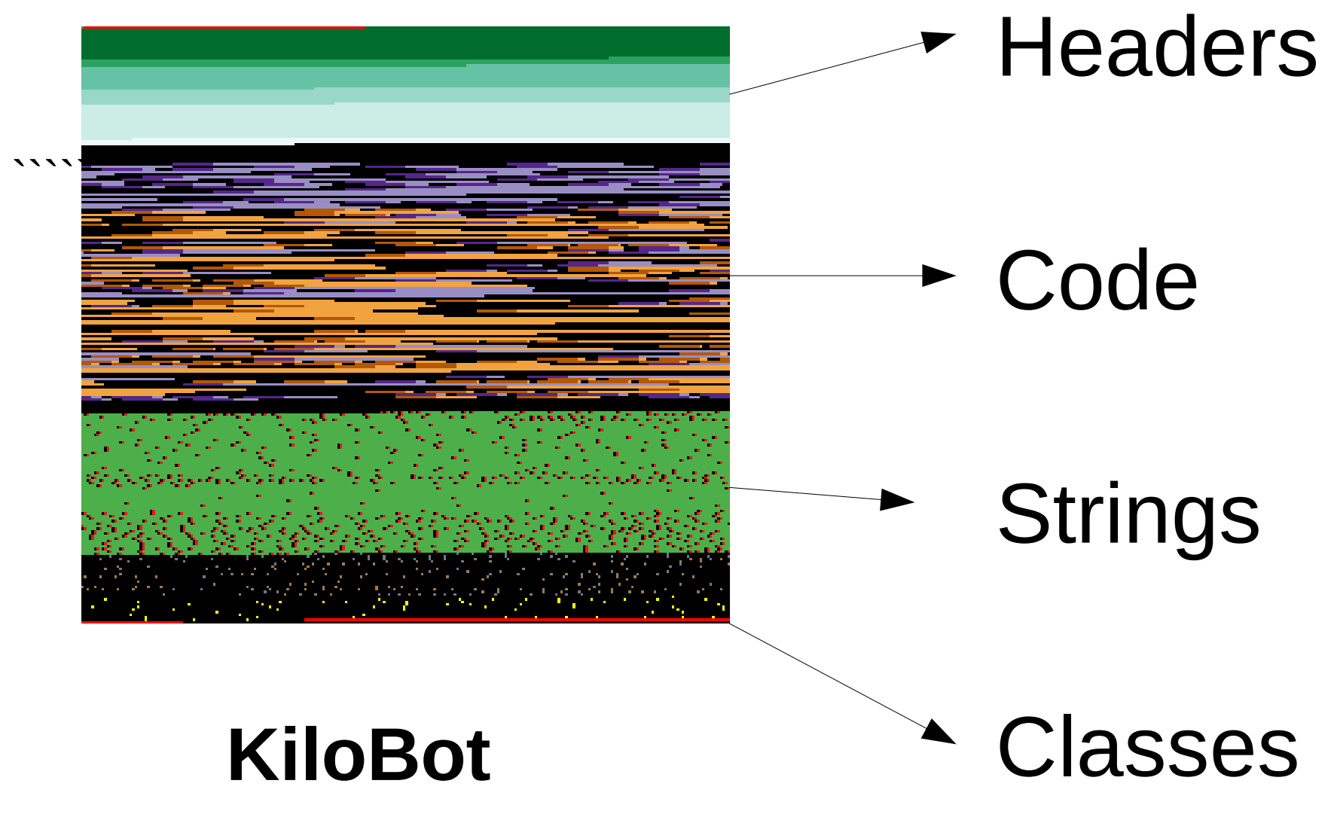
header	Structural information
string_ids	Offset list for strings
type_ids	Index list into the string_ids for types
proto_ids	Identifiers list of prototypes
field_ids	Identifiers list of fields
method_ids	Identifiers list of methods
class_defs	Structure list of classes
data	Code and data
link_data	Data in statically linked files



PikPok



Optimized version of Simplelocker

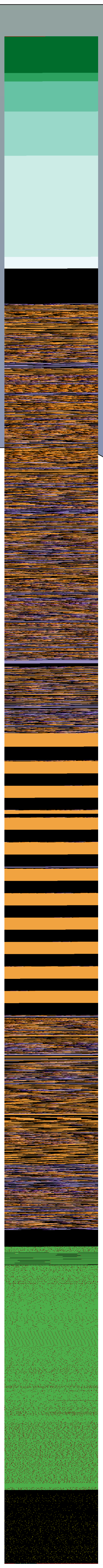


KiloBot

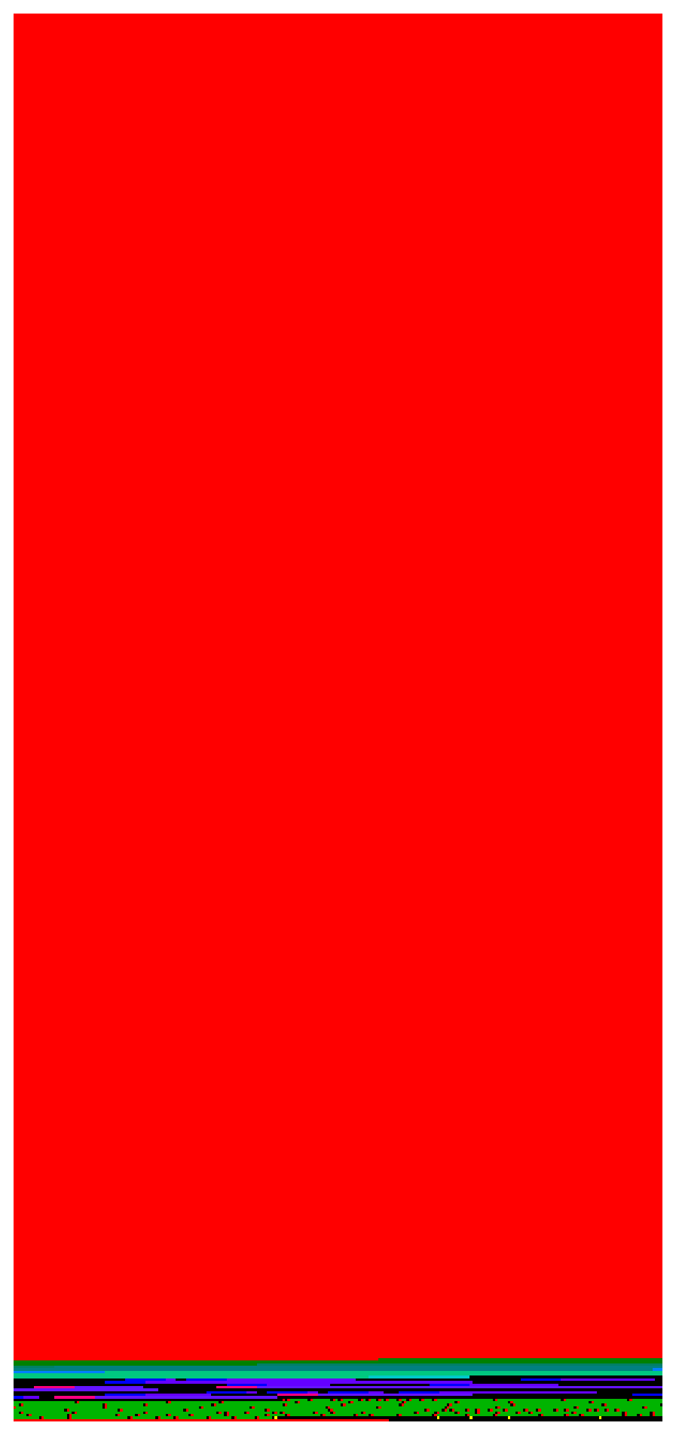
Game in Dot42

Regular tools

➢Normal header and sections.
➢Strings in order
➢Strings after code
➢Strings before class descriptors

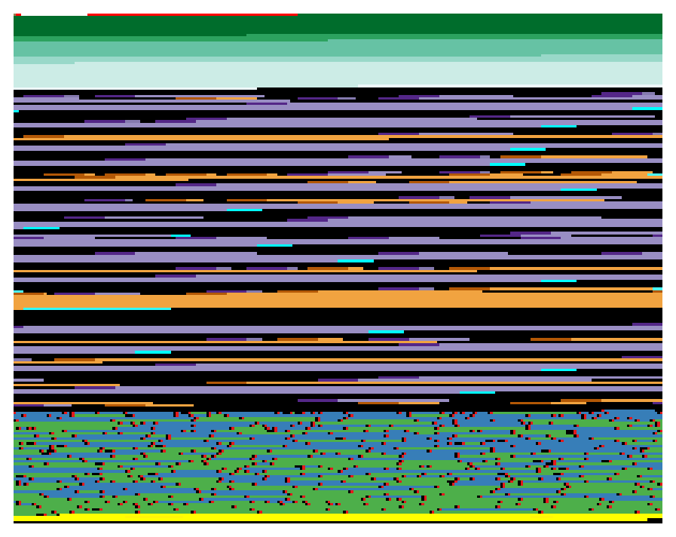


Generator Hello world



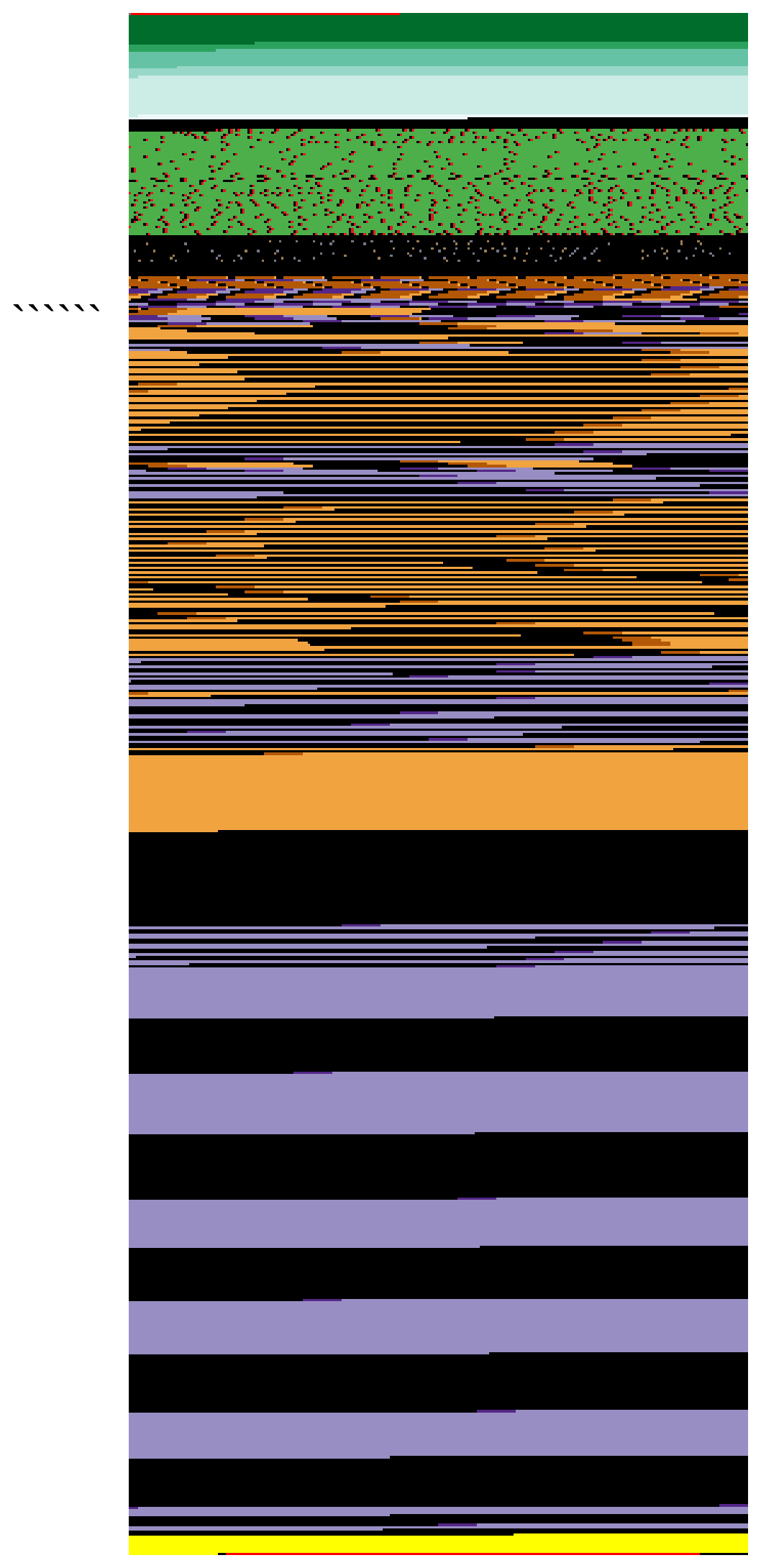
Huge header. Contains the original code

Hosed version of hello world



SendSMS

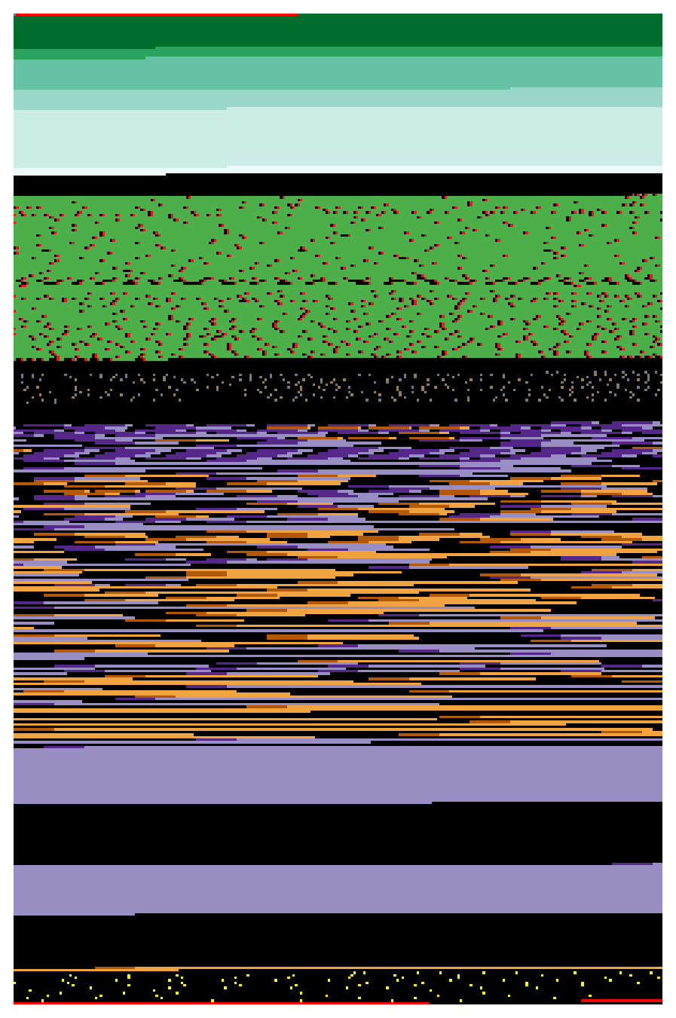
Strings not in order Produced by ApkTool



Dendroid

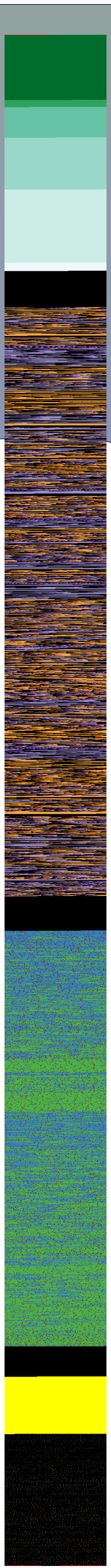
Strings before code. Code before classes

Produced by Proguard



Notifier

Other tools



Samsung Cisco

Work in Progress

➢Detecting the following tools :

- Android SDK (ADT with ant or Android Studio with gradle)
- Cross platform generators like Titanium, AdobeAIR, Phonegap
- Obfuscators like proguard or dexguard
- Packers or Encryptors like Hosed
- Third part compilers like Dot42 which convert .NET to dex
- Apktool commonly used to do reverse engineering or repackaging

Results and future work

➢The specification is straight forward, but there are some places where different tools behave different.

➢ To spot this differences we are using for now a visualization tool to create maps of the files and discover the differences. After manually analyzed this images, we will start the detection in an automatic way.