# SherlockDroid, an Inspector for Android Marketplaces

Axelle Apvrille - FortiGuard Labs, Fortinet
Ludovic Apvrille - Institut Mines-Telecom,
Telecom ParisTech, LTCI CNRS

Hack.Lu, Luxembourg
October 2014

FORTINET

TELECOM ParisTech

cnrs

# Who are we?

Axelle

Ludovic

# Many Android Applications

# Many Android Applications

# Unknown number of Android Apps

**We don't know exactly**
**how many apps there are**

# Unknown number of Android Apps

**We don't know exactly**
**how many apps there are**

- Precise number of Android marketplaces????
- How many duplicate apps?
- How many old/retired apps?

# Unknown number of Android Apps

**We don't know exactly**
**how many apps there are**

- Precise number of Android marketplaces????
- How many duplicate apps?
- How many old/retired apps?

**but it's BIG NUMBERS**

# Mobile Malware Infection Risk

**We don't know...** (exactly)

# Mobile Malware Infection Risk

**We don't know...** (exactly)

## What we **do** know

- Oct 2014. **840k** malicious Android **samples**

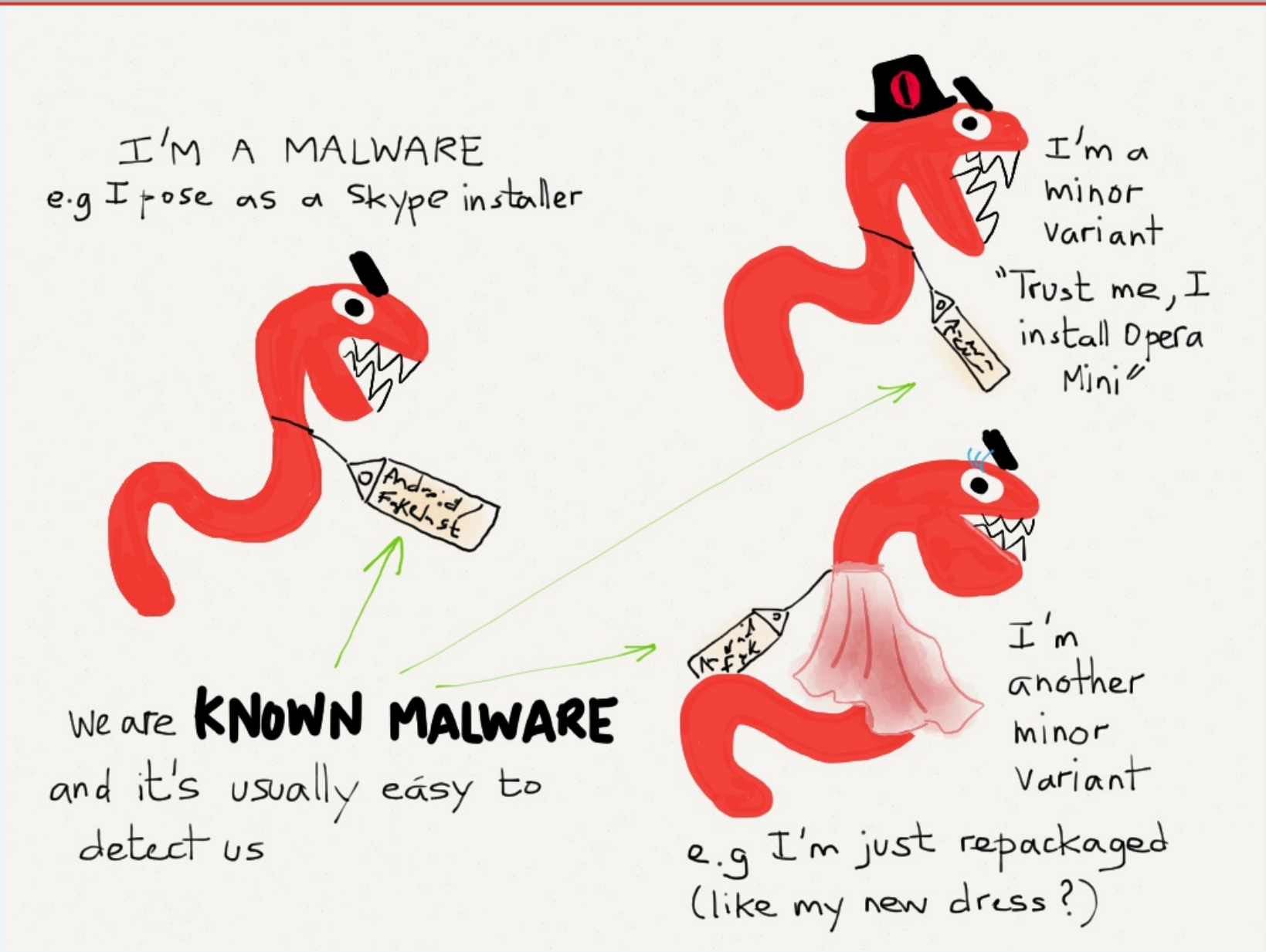
# Mobile Malware Infection Risk

**We don't know...** (exactly)

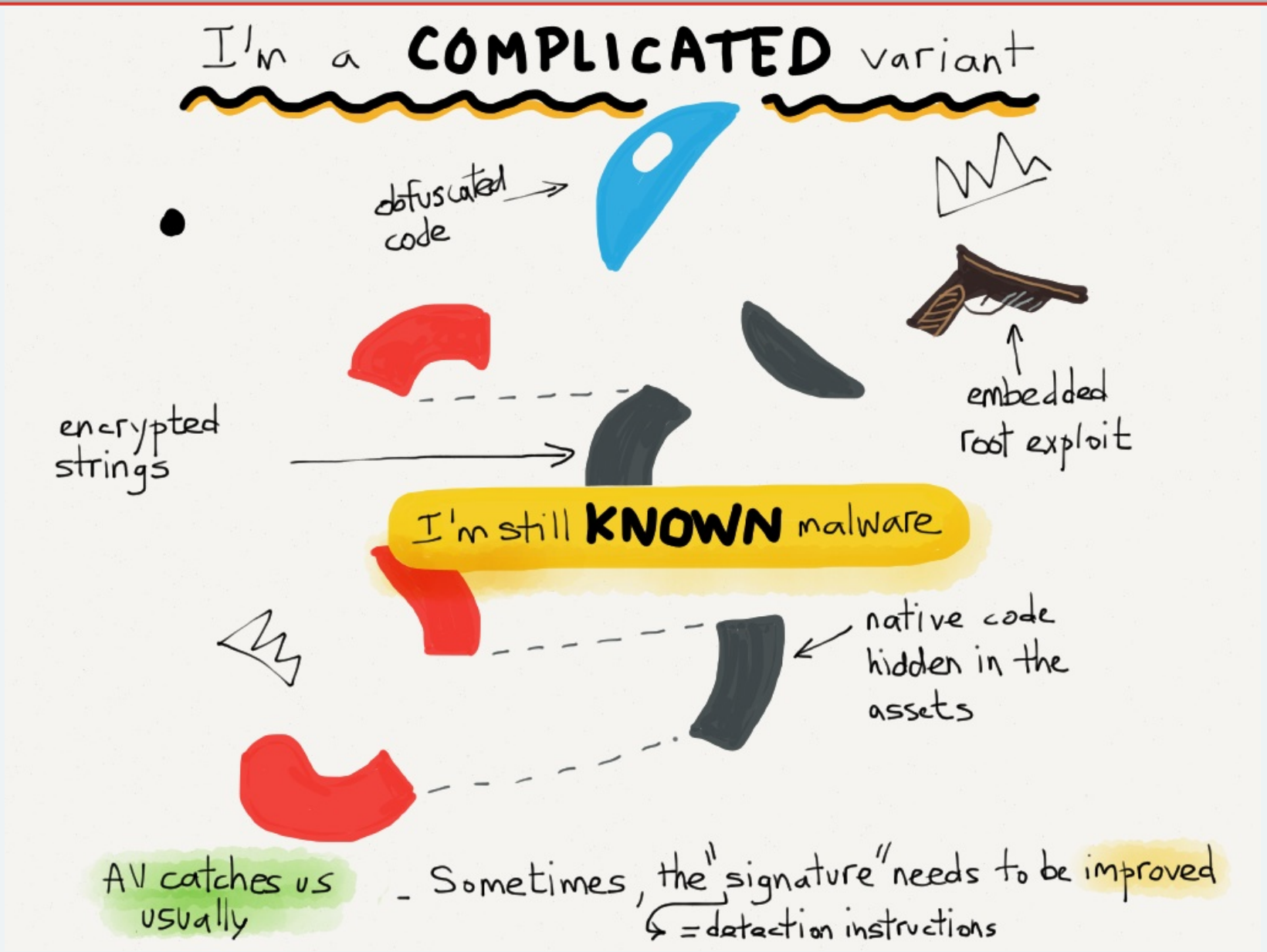
## What we **do** know

- Oct 2014. **840k** malicious Android **samples**
- **1,000+ new malicious Android sample every day**

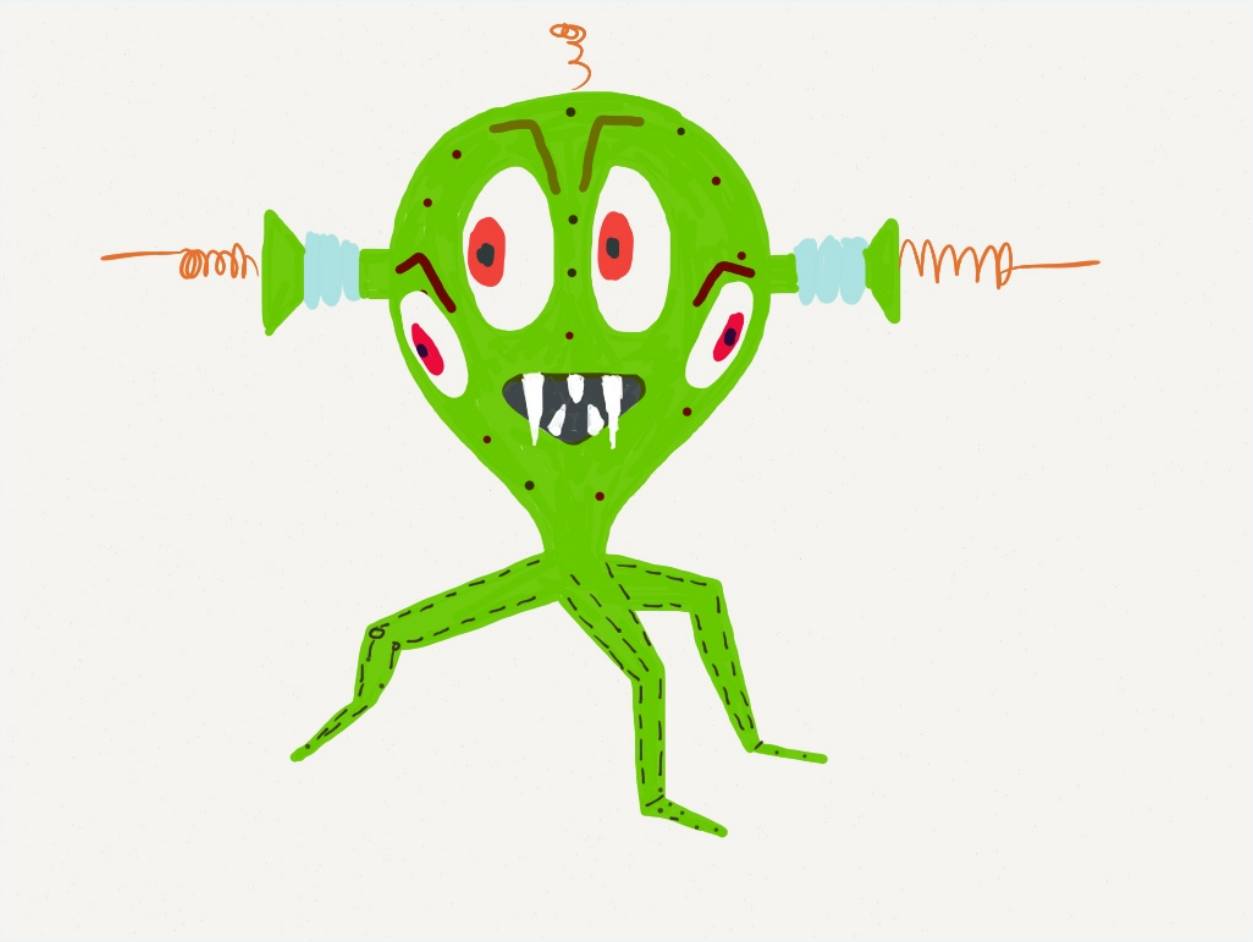
# Known Malware



# Known Malware

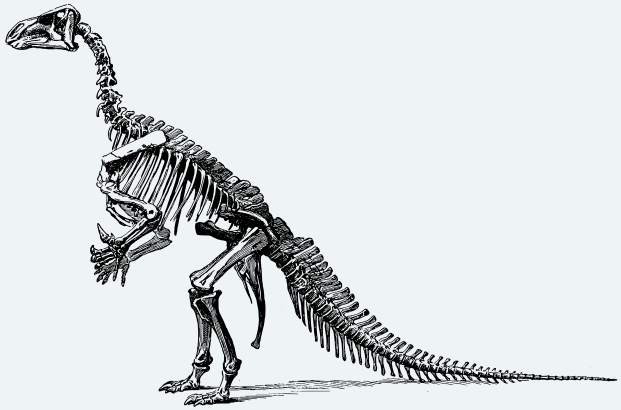


# Unknown Malware



**Do they exist? YES**

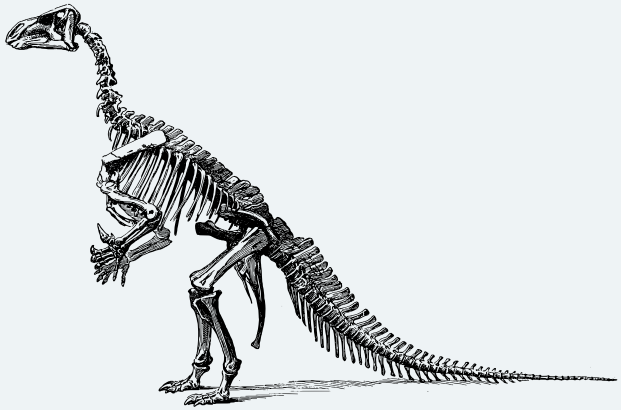
# Proof: Android Carbon 14 Dating ;)



Shortest detection delay for some samples by **all AV vendors**

| Name | Creation date | Detection date |
|---|---|---|
| Android/Wroba | June 16 | June 21 **+5d** |
| Android/Curesec | July 3 | July 11 **+8d** |
| Android/ScarePakage | July 13 | July 24 **+11d** |

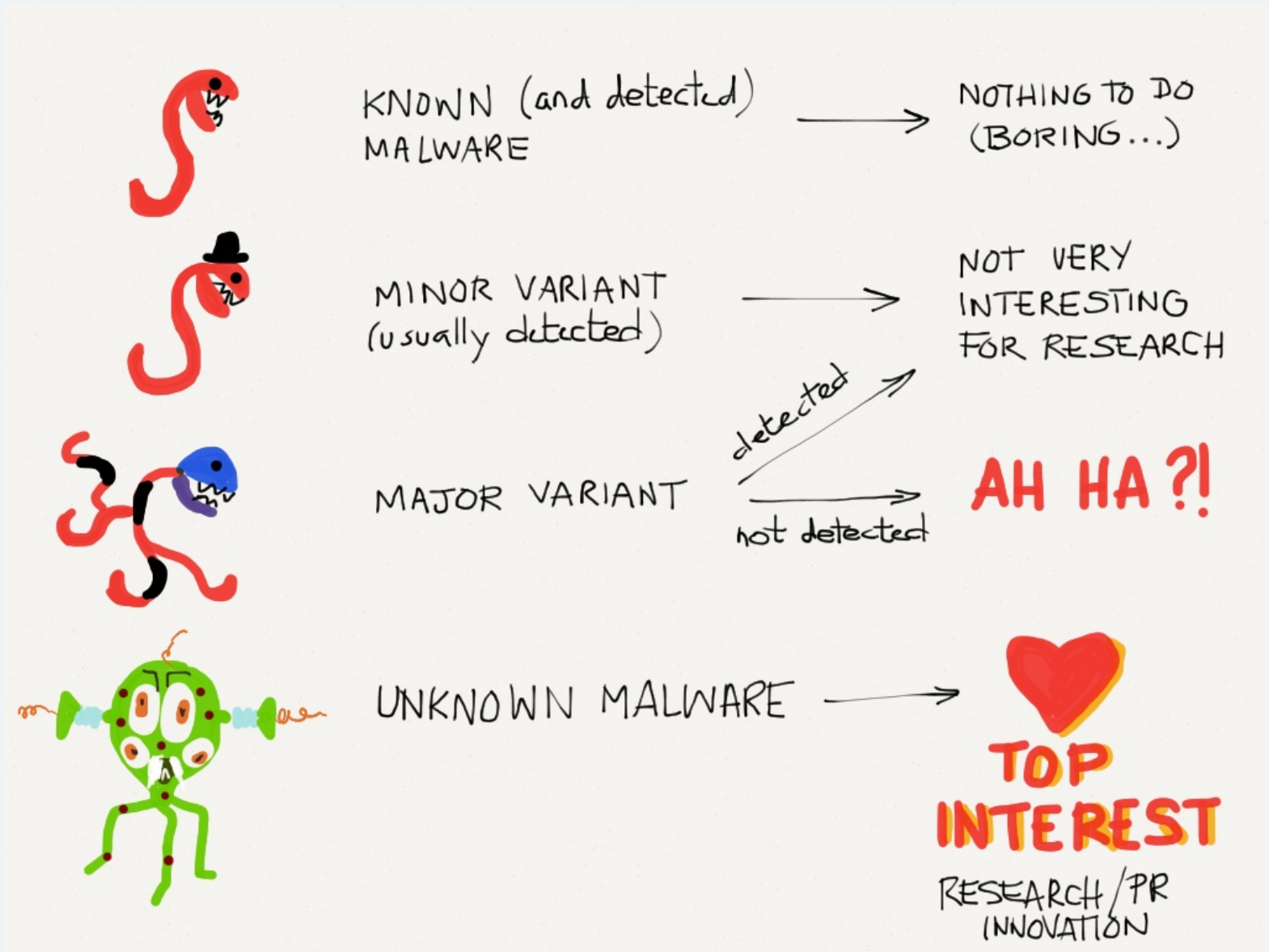
# Proof: Android Carbon 14 Dating ;)



Shortest detection delay for some samples by **all AV vendors**

| Name | Creation date | Detection date |
|---|---|---|
| Android/Wroba | June 16 | June 21 **+5d** |
| Android/Curesec | July 3 | July 11 **+8d** |
| Android/ScarePakage | July 13 | July 24 **+11d** |
| **Android/Ganlet** | **Nov 1 2013** | **May 15 2014 +6 months!!!** |

# So, What Are We Interested In?



# Problems with Manual Search

**Too many** apps and marketplaces to crawl

**Waste time** on clean apps

Even a **team of 100 analysts is insufficient**

# Problems with Manual Search

**Too many** apps and marketplaces to crawl

**Waste time** on clean apps

Even a **team of 100 analysts is insufficient**

**We need an automated system**

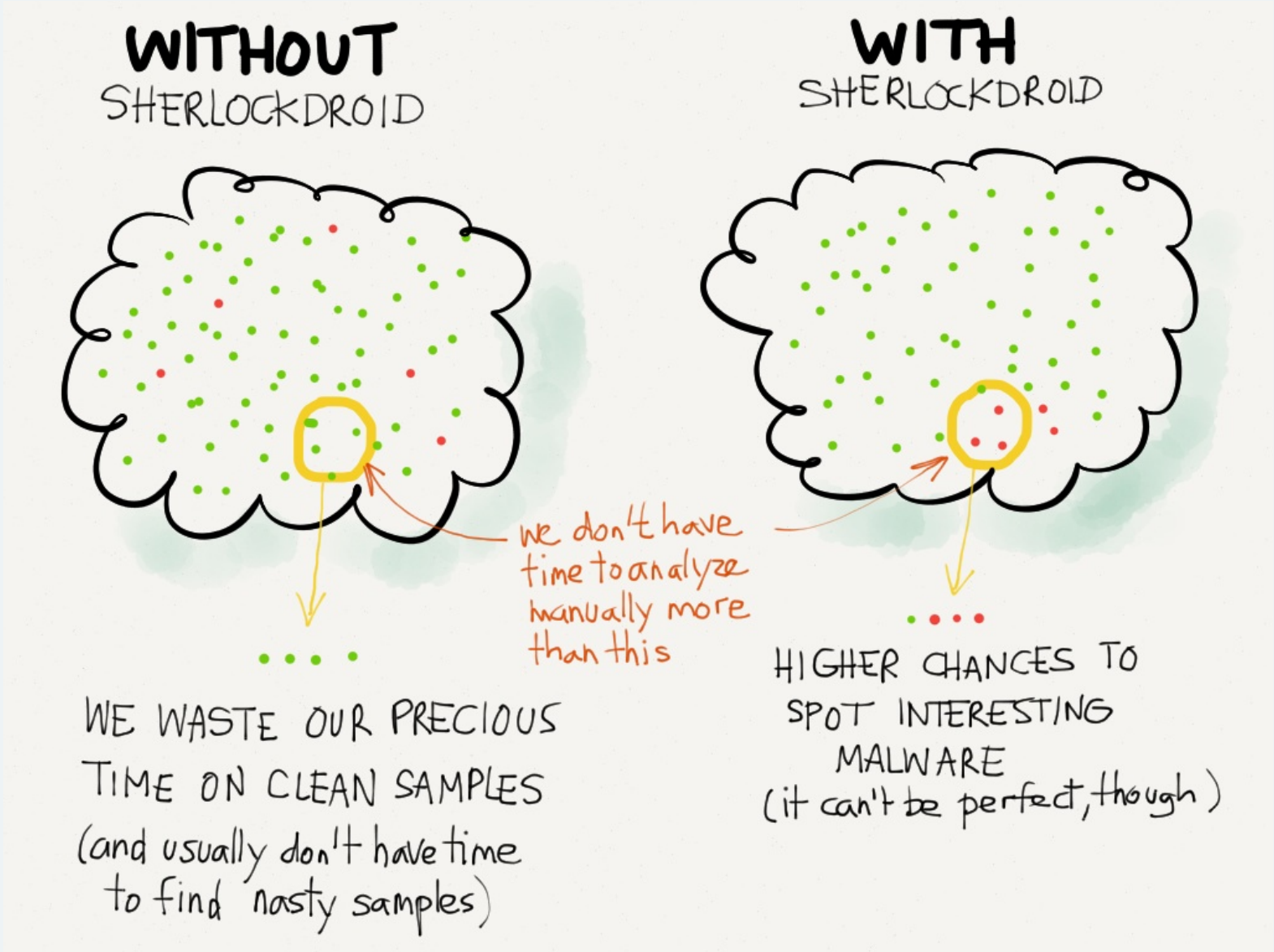
# SherlockDroid to the Rescue!

**Crawl Android marketplaces**

**Spot suspicious apps**

**Focus on major variants and unknown malware**

# SherlockDroid (Unbiaised) Benefits



# Remarks on SherlockDroid

**It is not an AV scanner**

because SherlockDroid does not handle known malware / minor variants

# Remarks on SherlockDroid

**It is not an AV scanner**

because SherlockDroid does not handle known malware / minor variants

**We will miss some malware**

we're not perfect :(

# Remarks on SherlockDroid

**It is not an AV scanner**

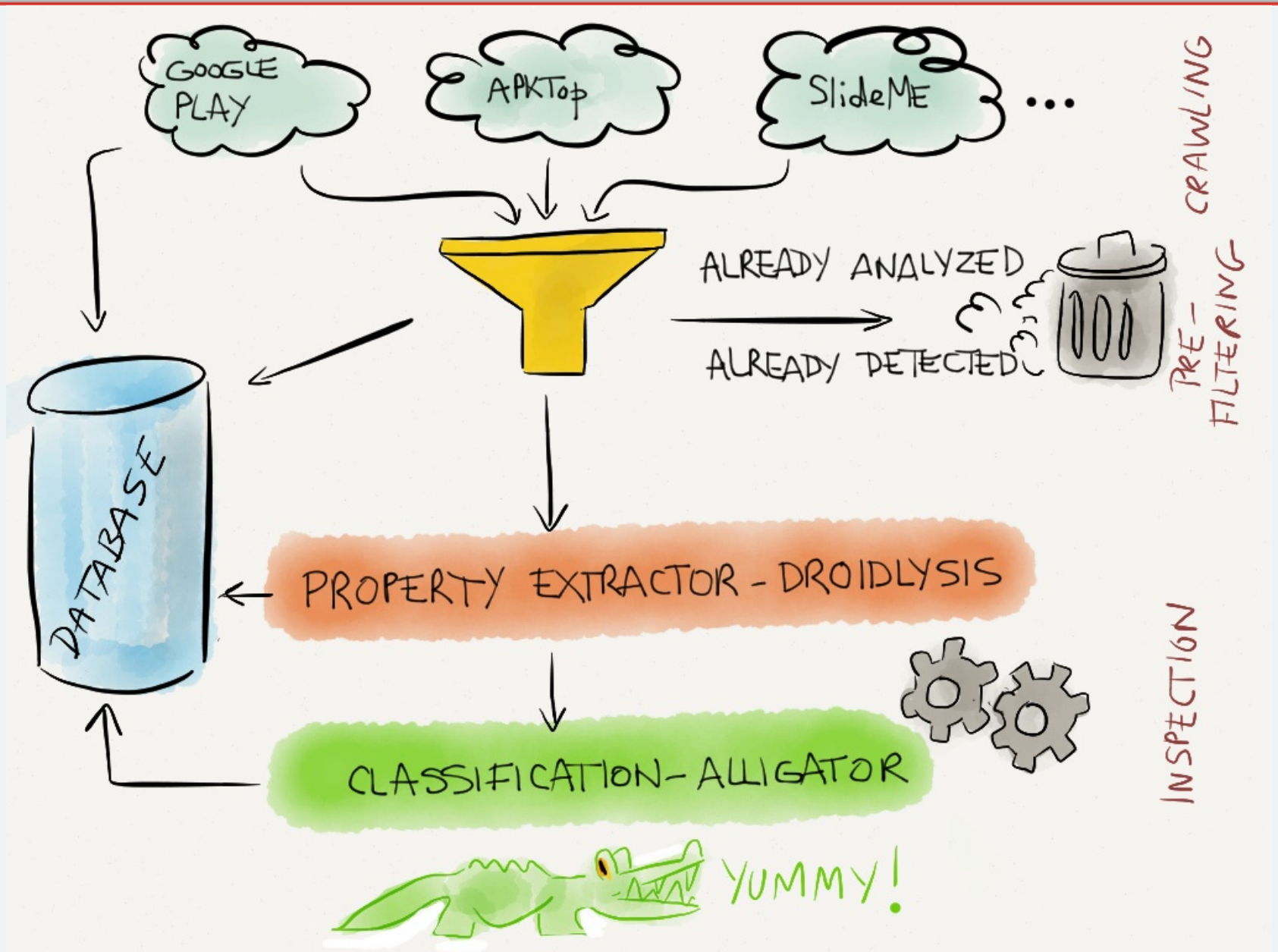
because SherlockDroid does not handle known malware / minor variants

**We will miss some malware**

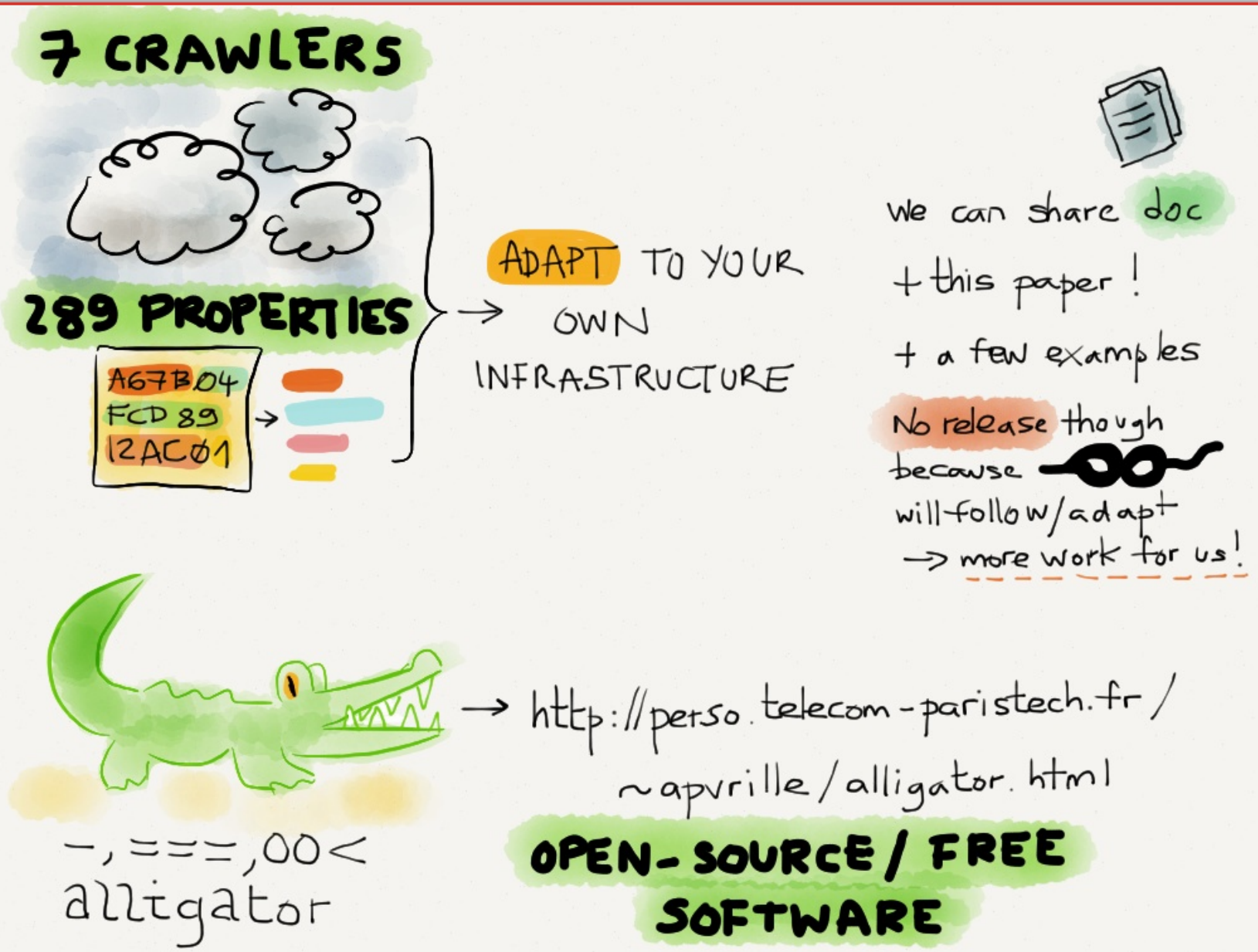
we're not perfect :(

*but we would have missed them without SherlockDroid too*

# SherlockDroid Architecture



# SherlockDroid: Current Status



# SherlockDroid: Current Tests

SherlockDroid is currently in 'heavy testing' phase

# SherlockDroid: Current Tests

SherlockDroid is currently in 'heavy testing' phase

**Crawled 140 K samples**

# SherlockDroid: Current Tests

SherlockDroid is currently in 'heavy testing' phase

**Crawled 140 K samples**

**Extracted properties of 550 K + samples**

# SherlockDroid: Current Tests

SherlockDroid is currently in 'heavy testing' phase

**Crawled 140 K samples**

**Extracted properties of 550 K + samples**

**Learning and classification: 480 K clusters!**
**At 50 K, FP: 0.99%, FN: 3.3%**

# SherlockDroid: Current Results

**8** new unknown malware
and Potentially Unwanted Apps

# SherlockDroid: Current Results

**8** new unknown malware
and Potentially Unwanted Apps
Okay, we would have preferred only nasty malware

# SherlockDroid: Current Results

**8** new unknown malware
and Potentially Unwanted Apps
Okay, we would have preferred only nasty malware

*Do you known any other framework who identified real unknown malware?*

# SherlockDroid: Current Results

**8** new unknown malware
and Potentially Unwanted Apps
Okay, we would have preferred only nasty malware

*Do you known any other framework who identified real unknown malware?*
**Answer:** DroidRanger: **2**

# SherlockDroid: Current Results

**8** new unknown malware
and Potentially Unwanted Apps
Okay, we would have preferred only nasty malware

*Do you known any other framework who identified real unknown malware?*
**Answer:** DroidRanger: **2**

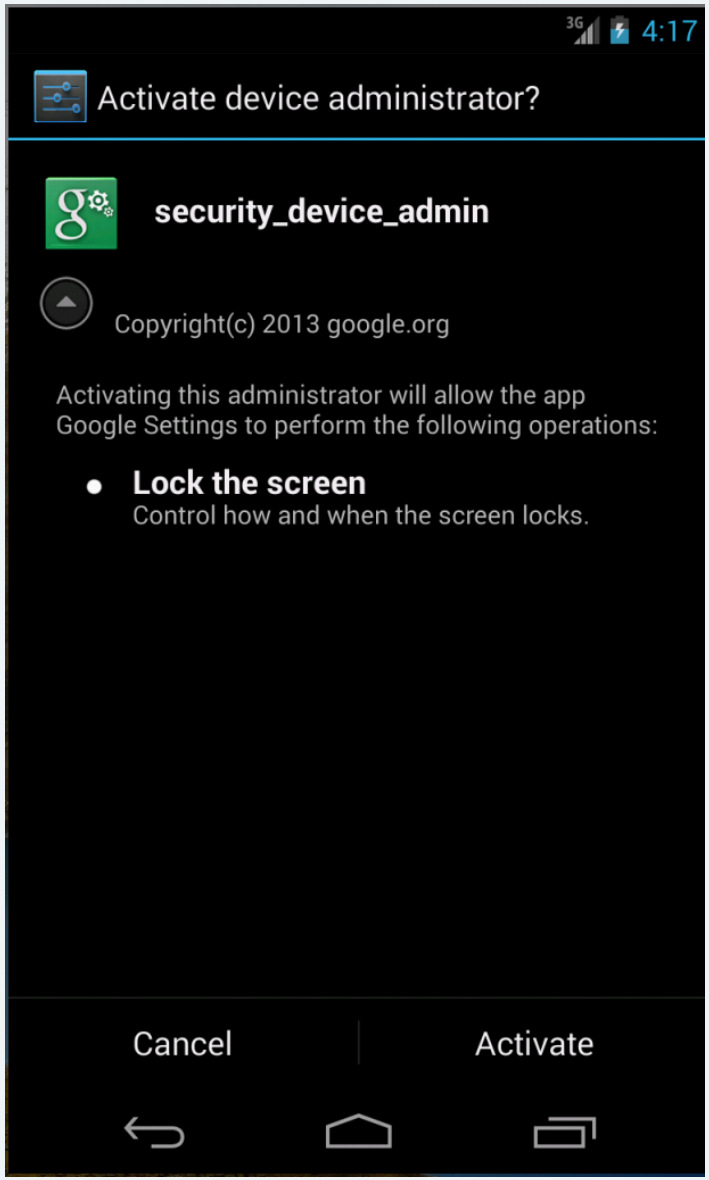
AAS, Andromaly, CopperDroid, Crowdroid, Drebin, MADAM, MAST, pBMDS, PUMA...
tested on *artificial* or *known malware*

# SherlockDroid: Hall of "Fame"

- Android/MisoSMS.A!tr.spy
- Android/Odpa.A!tr.spy
- Adware/Geyser!Android
- Riskware/Flexion!Android
- Riskware/SmsControlSpy!Android
- Riskware/Zdchial!Android
- Riskware/SmsCred!Android
- Riskware/Blued!Android

Descriptions: `http://www.fortiguard.com/encyclopedia/`

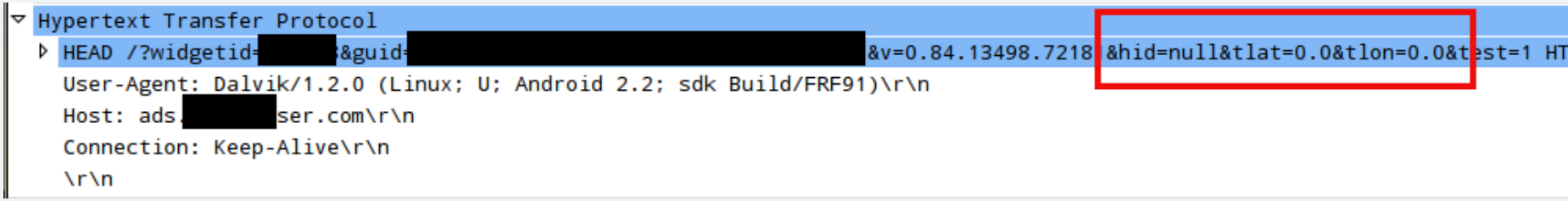
# Into Android/MisoSms Trojan Spyware




## Android/MisoSms.A!tr.spy

- Poses as Google Settings app
- Sends 1 initial email with phone number of victim
- Listens to incoming SMS
- Forwards them by email to attackers

# Into Geyser Adware




## Adware/Geyser!Android

Posts GPS location in clear text

http://blog.fortinet.com/post/alligator-detects-gps-leaking-adware

## LOL - In falsepositives.txt:

"Reputable companies including banks, US Government/ Military sector are using our tools"

# Crawlers - Evading Detection

**Easy to implement**
**but constantly needs to be maintained :(**

- Your IP: [redacted]
- URL: **www.appsapk.com/android/all-apps**
- Your Browser: libwww-perl/6.03
- Block ID: **BNP002**
- Block reason: Scanning tool access attempt.
- Time: Fri, 20 Jun 2014 05:30:21 -0400
- Server ID: **cp76**

- Search Limit
- Download activity per IP address
- User Agent verification
- Android ID verification `https://github.com/Akdeniz/google-play-crawler`

**Permissions are good ... but insufficient!**

# DroidLysis - Extracting Properties

**Permissions are good ... but insufficient!**

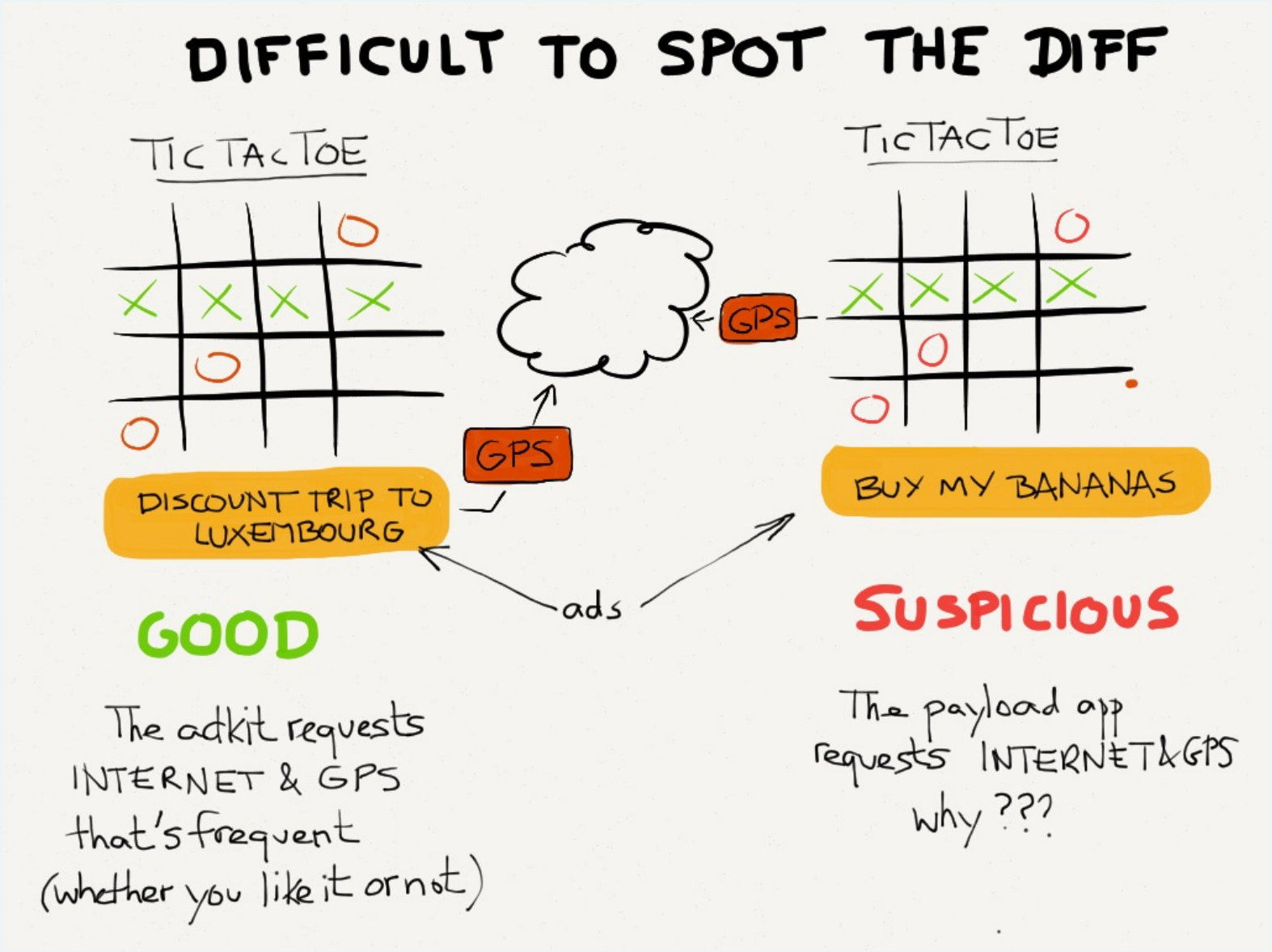
**In Dalvik, every object points to a string**

**Permissions are good ... but insufficient!**

**In Dalvik, every object points to a string**

**We also search assets and resources**

# Ruling out Third Party Code



## Gather clusters for learning - **only once**

Test with 1,514 newer clean samples
and 3,062 newer malware

| Learning cluster size | Learning time | Classification time time | FP | FN |
|---|---|---|---|---|
| 50,000 | 2 hours 13 min | 1 min 40 s | **0.99%** | **3.3%** |

We can favour minimum False Positives

## Gather clusters for learning - **only once**

Test with 1,514 newer clean samples
and 3,062 newer malware

| Learning cluster size | Learning time | Classification time time | FP | FN |
|---|---|---|---|---|
| 50,000 | 2 hours 13 min | 1 min 40 s | **0.99%** | **3.3%** |
| **480,000** | approx. 31 hours | 9 min 21 s | 1.8% | 0.5% |

It works with 480 K clusters !

We can favour minimum False Positives

## Gather clusters for learning - **only once**

Test with 1,514 newer clean samples
and 3,062 newer malware

| Learning cluster size | Learning time | Classification time time | FP | FN |
|---|---|---|---|---|
| 50,000 | 2 hours 13 min | 1 min 40 s | **0.99%** | **3.3%** |
| **480,000** | approx. 31 hours | 9 min 21 s | 1.8% | 0.5% |

SVM? Far worse! 50 K: **FP: 5.48%** FN: 0.65% !!!

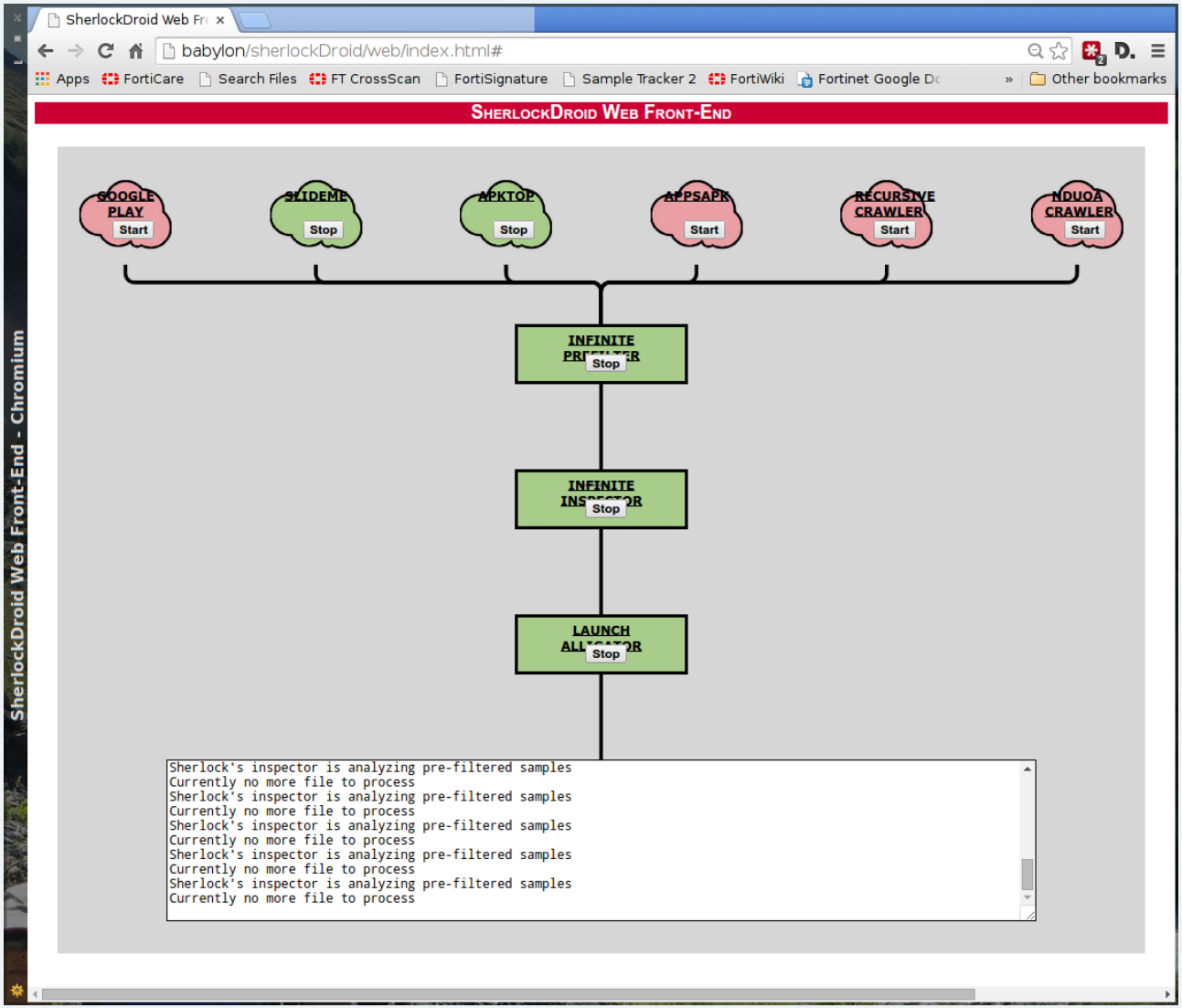
It works with 480 K clusters !

We can favour minimum False Positives

# Demo

Alligator unleashed!
Wake up!

# DEMO - SherlockDroid GUI [Preview]



# DEMO - SherlockDroid's Database

## Sample recently crawled, to pre-filter

```
115118|f8ef5f5306fb7...|net.mnprogram.mnagenda.apk|
Google Play|0|0|toanalyze||2014/10/09-14:04|967041
```

## Known malware

```
114902|ae084007fab965f829ba3fc...|
JJLord.30103.30000.visible.apk||0|0
|detected|Android/SMSreg.AK, SIGID: 49829716, VID: 5236396||0
```

## Unknown sample, to be inspected

```
115117|4dd15425c67b744125d7386...|
com.apalusa.lavoz.AgendaVos.apk|Google Play|0|0
|toanalyze||2014/10/09-14:04|2342643
```

## Unknown sample probably clean

```
115072|be849297862a50d7116d7a6be0...|
com.covertapps.joomlaadminmobilelite.apk|Google Play
|248.974979321754|145.030471289058|done||2014/10/09-13:48|583248
```

# DEMO - SherlockDroid Spots Suspicious Samples

## Example of suspicious samples

```
$ ./suspiciousApk.pl
suspiciousApk - show which samples are currently found suspicious by Al
Suspect: com.indvseng.indCENSORED.apk (f178c77d...
 origin: Google Play
 scoreRegular: 153.974979321754 scoreMalware: 161.923639714817
 difference: 7.94866039306393
------------
Suspect: floating-toucCENSORED.apk (3162b0c...
 origin: http://link.appsapk.com/downlo...
 scoreRegular: 153.974979321754 scoreMalware: 164.390159536531
 difference: 10.4151802147771
------------
Suspect: com.Ninjastrike456.ninjastrike.apk (65bb4...
 origin: Google Play
 scoreRegular: 153.621310611974 scoreMalware: 169.818181818182
 difference: 16.1968712062074
------------
Found 3 suspects
--- END
```

# DEMO - Cluster Sizes

## Size of clusters

```
$ wc -l learn-malware.csv learn-clean.csv
   guess-malware.csv guess-clean.csv
   486890 learn-malware.csv
    12368 learn-clean.csv
     3062 guess-malware.csv
     1514 guess-clean.csv
   503834 total
```

# DEMO - Example of Property File

## Size of clusters

```
105A663E.var,0.166667,0.000800,0,0.001930,0.000100,
0,0,0,0,0,0,0,0,1,0,0.201400,1,1,unknown,unknown,
0.000020,0,0.015000,0,0.000020,0.000010,0,0,0,0,0,1,
0,0,0,0,1,0,0,0,1,1,0,1,0,0,0,unknown,0,0,1,0,0,0,
0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0,0,1,
1,1,1,0,1,1,1,1,1,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,1,0,1,0,0,0,0,1,1,0,0,0,1,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,1,0,0,0,0,0,0.000010,0,0,0,0,0,0,0,0,
...
```

- Mostly boolean values (0, 1) + 'unknown'
- Integer values have been normalized to fit in [0,1]

# DEMO: Example of Learning Script

## The Alligator Language

```
setprintintermediatescore

printClusterSummary regular
printClusterSummary malware
printClusterSummary guess

setMultiplierWeight regular 0-100,1
setMultiplierWeight malware 0-100,1
compute inversedeviation

setPropertyWeightsFromColumn 63 6
...
setMultiplierWeight regular 0-100,1
setMultiplierWeight malware 0-100,1
compute inverseweightdeviation

setMultiplierWeight regular 0-100,1
setMultiplierWeight malware 0-100,1
compute degressiveproximity 5
...
```

# DEMO - Alligator Running

## Work! Work!

```
Alligator Daemon: AnaLyzing maLware wIth partitioninG and probAbiliTy-b
Rithms Daemon
 -,===,oo< Alligator: (C) Institut Mines Telecom / Telecom ParisTech, L
VRILLE, ludovic.apvrille@telecom-paristech.fr
 -,===,oo< http://perso.telecom-paristech.fr/~apvrille/alligator.html
 -,===,oo< Alligator is released under a CECILL License. See http://www
nfo/index.en.html
 -,===,oo< Enjoy!!!

*** Your Alligator version is: 0.3-beta1 -- build: 1433 date: 2014/10/0
4 CET ***

-,===,oo<  1/7   0% unknown | 85MB/1820MB
-,===,oo<  2/7   0% unknown | 97MB/1820MB
-,===,oo=  2/7   0% unknown | 97MB/1820MB
-,===,oo<  2/7   0% unknown | 97MB/1820MB
-,===,oo=  2/7   0% unknown | 97MB/1820MB
...
```

# DEMO - Alligator Report

## Classifying samples

```
*** Overall report of guess ***
Classification time:468.121s

** Overall results **
regular - 11249 elements in cluster, nb of properties: 288
malware - 50000 elements in cluster, nb of properties: 288
guess - 3 elements in cluster, nb of properties: 288

Results summary:
2 regular(s) found, 1 malware(s) found in guess
Percentage of regular: 66.66666666666666
Percentage of malware: 33.33333333333333
regular: Light:2 (66.67%)  Medium:0 (0.00%)  Strong:0 (0.00%)
malware: Light:1 (33.33%)  Medium:0 (0.00%)  Strong:0 (0.00%)
105A663E.var:  regular  (regular:131.36352883261992, malware:121.909090
...
```

# Thank You

## Contact info

SherlockDroid: aapvrille at fortinet dot com
Alligator: ludovic dot apvrille at telecom minus paristech dot com

## Downloads

Alligator Release

L. Apvrille, A. Apvrille, *Pre-filtering Mobile Malware with Heuristic Techniques*, GreHack 2013

A. Apvrille, T. Strazzere, *Reducing the Window of Opportunity for Android Malware*, EICAR 2012

Powerpoint slides? No way! This is LaTeX- Beamer !