Evaluating the signature based and research antimalware tools against malware in the wild and third-party markets: A technical report

Francesco Mercaldo*

Corrado Aaron Visaggio**

Assunta Oropallo

Paolo Pirone

**fmercaldo@unisannio.it*

visaggio@unisannio.it* *(contact author)

April 2015

Technical Report

Corso Garibaldi 107, Benevento - Italy

Abstract

Android malware is spreading more and more. But the current solutions are able to detect the malware on our device? In the following report we analyze the effectiveness of 22 signature based antimalware and two research prototypes (Andrubis and Androguard), through submission of a dataset of 5560 malicious apps.

We also collected another dataset of 4000 applications from two third-party markets, AppChina and Gfan, with the aim to find out if the alternative markets are hiding applications with malicious purposes.

The evaluated antimalware

Andrubis

Andrubis[1] represents one of the most well-known research prototypes in the field ofmalware detector; it was developed by the International Secure Systems Lab and is a service that has been operating since June of 2012. In just two years of activity have been analyzed more than one million Android application. Andrubis is an extension of the existing service instead Anubis specializing in Windows Malware Analysis. It is possible to access the features offered by Andrubis: either through the browser by connecting to the link http://anubis.iseclab.org/, or through a script for automatic submission or through an app dedicated.

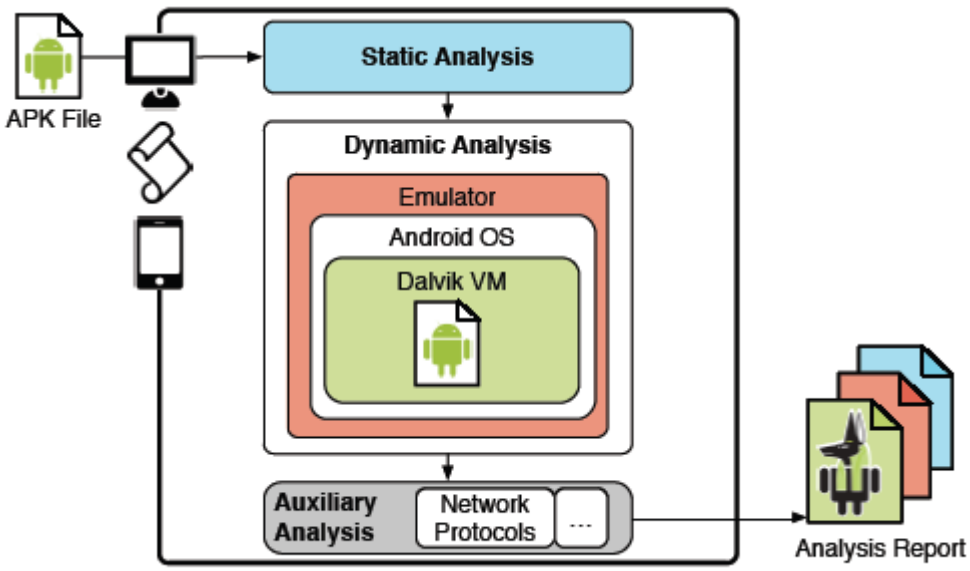

Figura 1: Andrubis system overview

Figure 1 shows Andrubis system overview, basically it uses an hybrid approach based on static, dynamic and auxiliary (this analysis is referring to network protocols) analysis.

Androguard

Androguard[2] is a tool written in python, used to perform various operations and analysis of Android applications such as:

• .dex/.apk file manipulations;

• operations like disassemble / decompilation of .apk/.dex files;

• static analysis of code;

• assess the diffing (differences) Android app;

• determine if an application contains malicious code.

The use that will be made in this paper will be mainly focused on the malware-detection, i.e. on the ability of the tool to detect malicious code in Android applications.

Jotty

To evaluate simultaneously the most number of antimalware with the signatures update as a possible, we use the Jotti[3] malware scan. Jotti is a free service that offers the scan of a candidate files from more than 20 antimalware, it offers also a desktop client to submit candidate applications without the necessity to use the browser to upload every single application. It uses the Linux version of the antimalware. It has only a limit: the size of the upload cannot exceed 25MB, but the samples in our dataset do not beat this limitation. If an application is marked as a malware, Jotti retrieves also the family it belong to (for each antimalware), as opposite the application is marked as a trusted one.

Table 1 shows the antimalware available using the Jotti service:

AntiMalware	Web site
Ad-Aware	http://it.lavasoft.com/
Agnitum	http://www.agnitum.com/
Arcabit	http://www.arcabit.pl/
Avast	https://www.avast.com/it-it/index
Avg	http://www.avg.com/it-it/homepage
AntiVir	https://www.avira.com
Bitdefender	http://www.bitdefender.it/
Clam AV	http://www.clamav.net/index.html

Dr.Web	http://www.freedrweb.com/cureit/?lng=it
eScan	http://www.escanantivirus.it/
Eset	https://www.eset.it/
Fortinet	http://www.fortinet.com/
F-Prot	http://www.f-prot.com/
F-Secure	https://www.f-secure.com/
GData	https://www.gdata.it/
Ikarus	http://www.ikarussecurity.com/
Kaspersky	http://store.kaspersky.it/
Panda	http://www.pandasecurity.com/
Quick Heal	http://www.quick-heal.it/
Sophos	http://www.sophos.com/
Trend Micro	http://www.trendmicro.it/
VBA32	http://www.anti-virus.by/en/vba32arkit.shtml

Table 1: the 22 antimalware used to evaluate the signature-based detection

The Experiment

The aim of the experiment is to evaluate the effectiveness of current free and commercial antimalware solutions and of two research prototypes, i.e. Andrubis and Androguard.

The report poses the following research questions:

- **RQ1**: given a set of malware mobile applications how is effective in their detection Andrubis?
- **RQ2**: given a set of malware mobile applications how is effective in their detection Androguard?
- **RQ3**: given a set of malware mobile applications how are effective the current signature-based antimalware in their detection? Are they able to classify the samples in the family they belong to?
- **RQ4**: given a set of application downloaded from third-part markets, how are considered as malware from the Andrubis point of view?
- **RQ5**: given a set of application downloaded from third-part markets, how are considered as malware from the Androguard point of view?
- **RQ6**: given a set of application downloaded from third-part markets, how are considered as malware from the current signature-based antimalware?

As research questions have explained we use two dataset in order to response them: a first dataset containing 5560 malware samples, classified in the families they belong to, and a second dataset, containing applications of untrusted sources, i.e. potentially malware.

The malware dataset is released by research community[4,5] as Drebin project, it contains 5560 classified in 179 families, it represent the most populous and more recent mobile malware application repository, the samples was categorized from Agoust 2010 and October 2012. We listed in table 2 the top 20 populous families in the dataset sorted by number of samples:

Family	#samples
FakeInstaller	925
DroidKungFu	667
Plankton	625
Opfake	613
GinMaster	339
BaseBridge	330
Iconosys	152
Kmin	147
FakeDoc	132
Geinimi	92
Adrd	91
DroidDream	81
ExploitLinuxLotoor	70
MobileTx	69
Glodream	69
FakeRun	61
SendPay	59
Gappusin	58
Imlog	43
SMSreg	41

Table 2: the top 20 populous family in the mobile malware dataset.

The second dataset was retrieved using a python crawler developed by authors of the report. We have selected the top 2 most used third-party markets, AppChina (http://www.appchina.com/) and Gfan (http://www.gfan.com/), and we run the crawler to download a total of 4000 applications, 2000 from AppChina and 2000 from GFan.

We submitted a total of 9560 (5560 malware + 4000 unknown) applications to 22 antimalware product and to Andrubis and Androguard.

Evaluating the malware dataset

In this section we discuss the results deriving from the analysis of the malware dataset using Jotti service, Andrubis and Androguard. Regarding Andrubis, we point out that Andrubis was able to analyze 5169 malware on the full dataset composed by 5560 of them. The reason why the remaining samples was not analyzed is that Andrubis has an 8 MB limitations of file size upload and in several cases it was due to offline emulator. Androguard has no size file uploading limitation.

In following graphs we explain the results. Andrubis assigns a maliciousness rank to scanned applications: it ranges from 0 to 10. Since Andrubis does not provide a threshold score beyond which an application can be classified as malware, we consider the rank values over 6 indicative of a malware from. We also divided the possible value range in 4 interval, because we think that there is a difference between an application classified with 0 than another one classified with 5 in terms of maliciousness. Table 3 shows the malicious classes we use to discriminate a malware from a trusted applications:

Maliciousness category	**score**
Trusted	0≤score≤1
probably trusted	1≤score≤6
probably malware	6≤score≤9
malware	9≤score≤10

Table 3: the maliciousness category for Andrubis rank values

Anyway, all the applications resulting with a rank over 6 was considered as a malware, in contrast with the applications with a rank under this threshold was identified as a trusted samples.

The graph in following figure (fig. 2) shows the percentage of the dataset malware samples that have returned in the categories following explained.

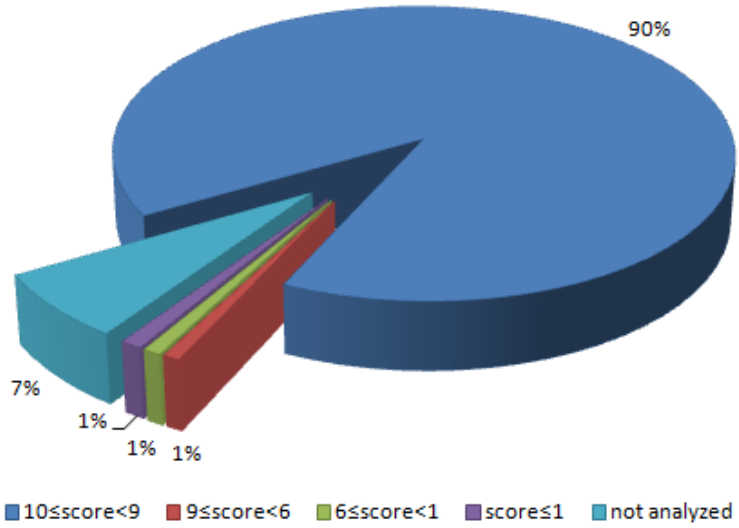

Figure 2: population percentage of malware samples categories

The figure shows that 91% of malware samples are correctly recognized as malware (5022/5576) in the fourth category (9≤score≤10), the probably malicious category reach a percentage of 1% with only 45 samples belonging the this one. Regarding the last two categories, only 1% of samples were ranked as a probably trusted (46/5560) and another 1% as trusted (56/5560). 391 applications, i.e. the 7% of the dataset was not analyzed (391/5560).

To 817 applications in the fourth category (malware category, with 9≤score≤10) has been assigned a rank equal to 10, the maximum maliciousness rank provided by Andrubis.

The following graph focalize this point: it shows the effective value obtained and in red are highlighted the applications resulting with a score equal to 10.0.

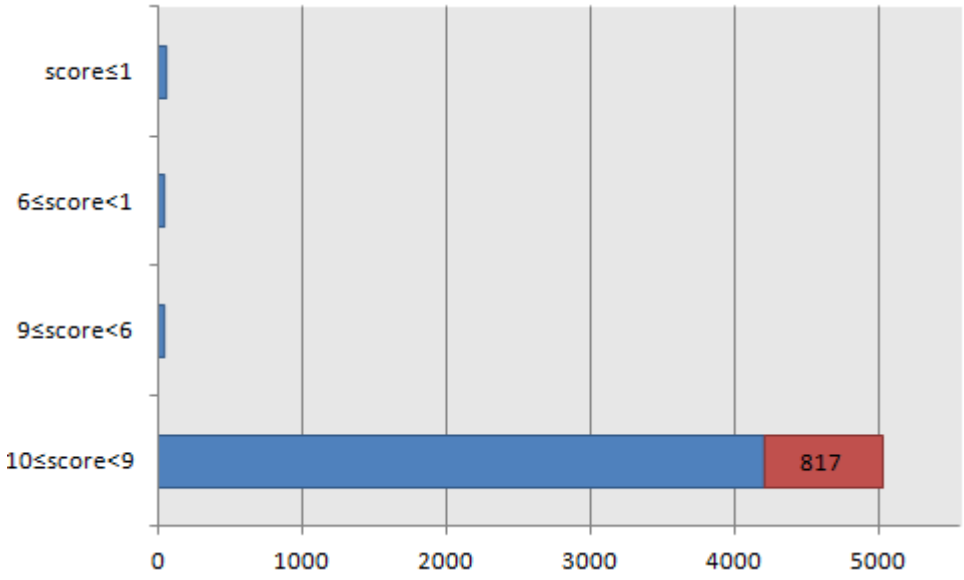

Figura 3: applications ranked with the maximum value (10.0) from Andrubis.

From the results there is evidence that Andrubis right classified the majority of submitted malware. As matter of fact, the true positive percentage, i.e. malware correctly classified, is equal to 98% and the false negatives percentage, i.e. malware classified as trusted is represented by only the 2% (neglecting the 7% that was not analyzed). Figure 4 shows this result.

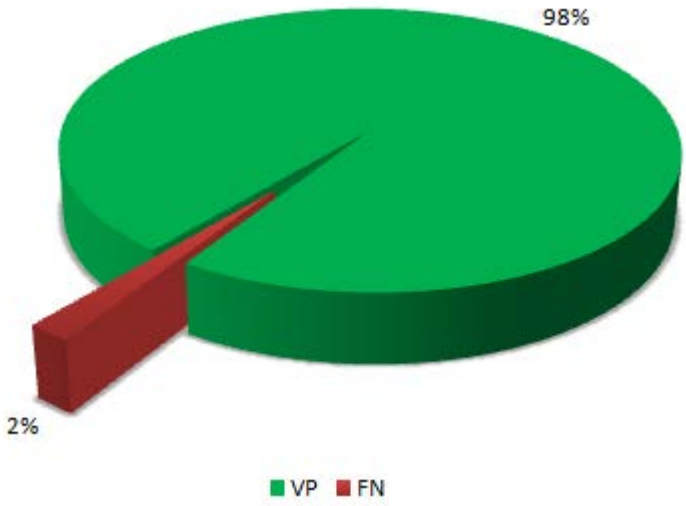

Figure 4: true positive (TP) and false negative (FN) percentage

The report delivered by Andrubis does not provide any information about the family belonging the sample indentified as malware, for this reason we can not analyse other metrics like, for instance, the malware family most recognized and so on. But, we conduct a supplementary analysis to discovery how malware family samples are in the false negative percentage, i.e. how family were unrecognized by Andrubis?

Figure 5 shows all the family containing at least one samples recognized as trusted from Andrubis. We express percentage ratio between samples classified as trusted and the number of samples belonging to the family considered:

$$r\% = \frac{samples\ classified\ as\ trusted\ in\ family\ x}{total\ samples\ belonging\ to\ family\ x} * 100$$

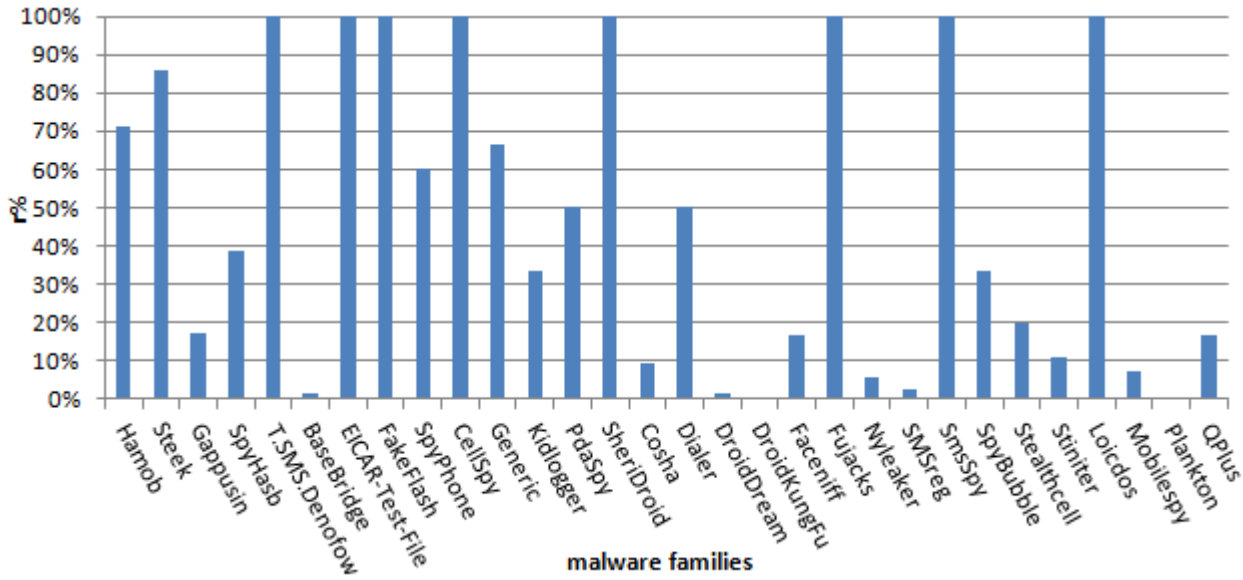

Figure 5: malware families recognized as trusted histograms

As result, the less recognized families are *Hamob*, *Steek* and *Gappusin*. In addiction we observe that several families presents a 100% rate, but the majority of these families present a very few number of samples, in some cases also only one.

RQ1 response: only the 2% of the malware dataset was classified from Andrubis with a rank less than 6 (i.e. as trusted), we conclude that all samples belonging the most famous and diffused families were correctly detected from Andrubis.

To response to RQ2 we submitted the malware dataset to Androguard, the second research prototype we tested in this report.

The following graph compares the number of malware samples divided into families present in the dataset and the number of samples correctly identified as malware and classified into families belonging to 20 families most populous.

To be noted that *FakeInstaller* the larger family in our dataset has never been detected by Androguard.

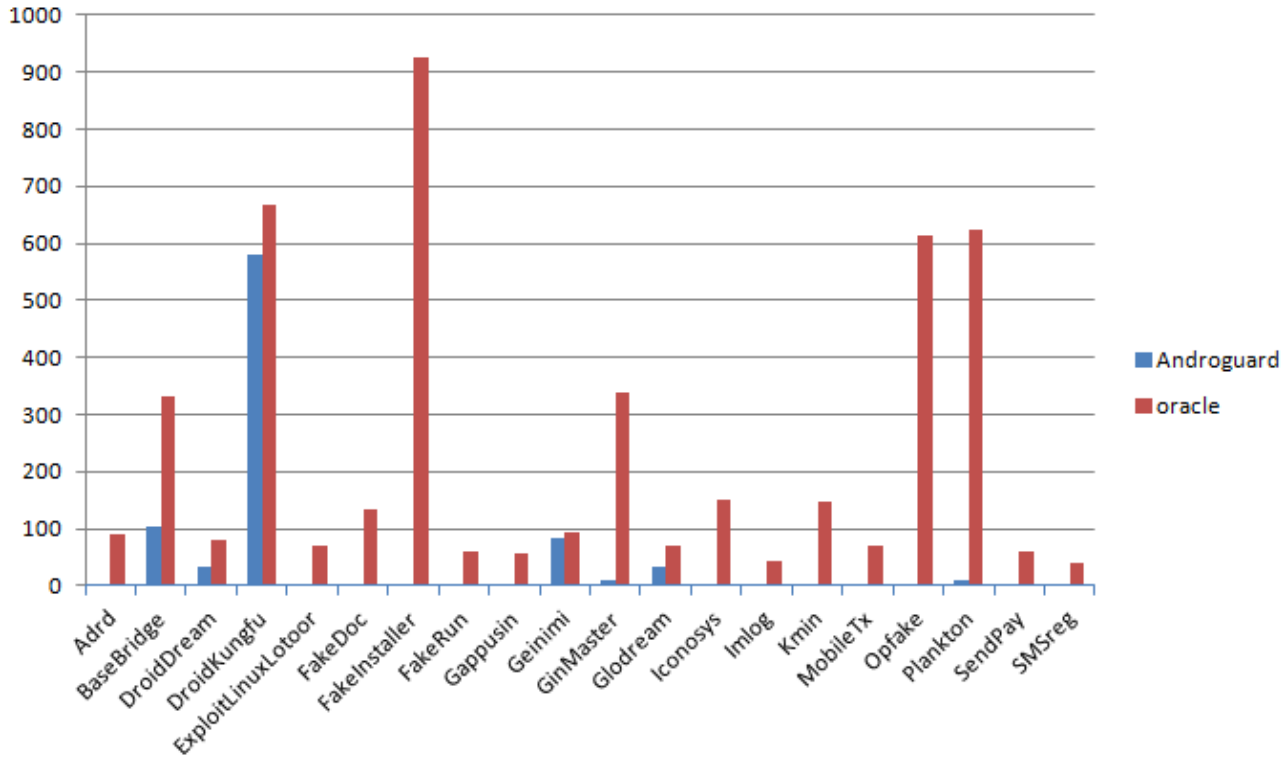

Figure 6: comparison between families in our dataset and the number of samples reconized in the families by Androguard

We obtain a true positive equal to 0.22, i.e. 22% of submitted samples was rightly classified as malware, while regarding false positive rate we obtain a percentage equal to 0.7, i.e. 78% of submitted samples were mistake as trusted.

Following histograms explain our results detailed for each malware family analyzed.

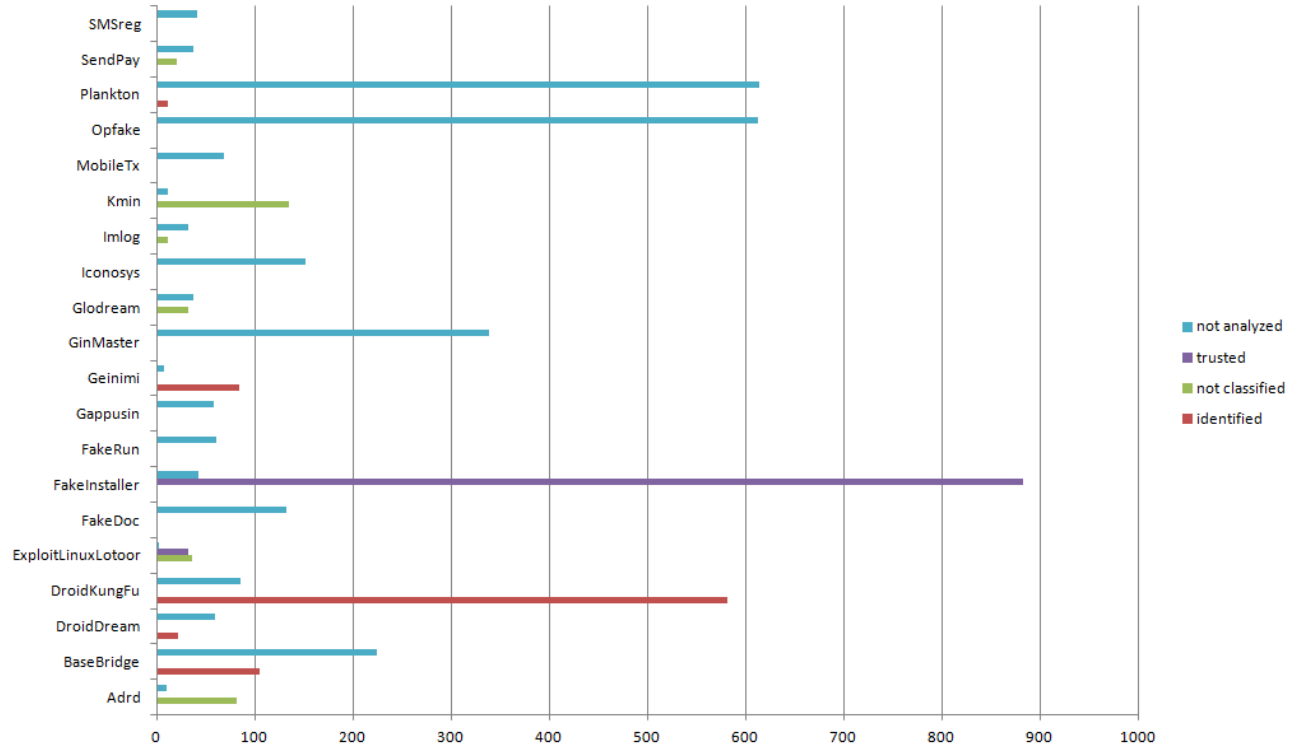

Figure 7: histograms detailing the malware family detection in Androguard.

DroidKungFu is the most classified malware, 581 samples out of 667 were associated with the right malware family. *Fakeinstaller* samples, as also evidenced, were recognized as trusted. Finally we note how many samples, especially those who belong to the family of *Plankton*, *Opfake* and *GinMaster*, have not been analyzed, i.e. they were not just scanned by Androguard.

We summarize the Androguard analysis in the following figure.

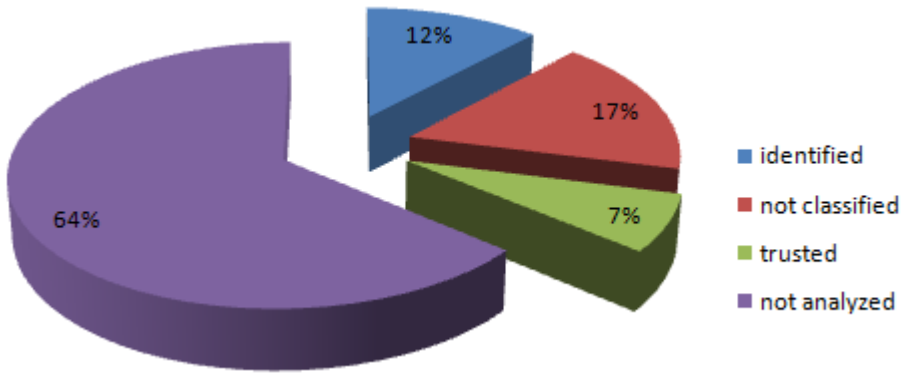

Figure 8: percentage average regarding samples identified, classified in the wrong family and as trusted and not analyzed by Androguard.

Androguard has detected the right malware family in 12% of samples, 17% was also detected as malware but not classified in the right family, 7% was recognized as trusted, while 64% was not analysed by Androguard.

RQ2 response: Androguard exhibits a percentage equal to 22% in malware detection, Regarding the classification it recognized the right family with a percentage equal to 12%. 17% of malware samples was Identified as trusted.

To response to RQ3 we consider the antimalware provided by Jotti service. In this case all samples were analyzed, the size limit for upload was 25MB.

The following table explains the results we obtained. We report for each antimalware the number of samples correctly detected as malware (we recall that the number of samples in the malware dataset is 5560), the false negative number, i.e. the ratio from the applications detected as trusted and the number of application in the dataset (5560) and the true positive value as percentage.

Antivirus	***Malware detected***	***False Negative***	***True Positive***	***TP percentage***
Ad-Aware	5436	0.022302158	0.977697842	98%
Fortinet	5161	0.07176259	0.92823741	93%
Agnitium	465	0.916366906	0.083633094	8%
F-PROT	836	0.849640288	0.150359712	15%
Arcabit	5433	0.022841727	0.977158273	98%
F-Secure	5453	0.019244604	0.980755396	98%
Avast	3751	0.325359712	0.674640288	67%
G Data	3081	0.445863309	0.554136691	55%
AVG	3081	0.445863309	0.554136691	55%
Ikarus	4665	0.160971223	0.839028777	84%
Antivir	5432	0.023021583	0.976978417	98%
Kasperky	5328	0.041726619	0.958273381	96%
Bitdefender	5321	0.042985612	0.957014388	96%
Panda	5154	0.073021583	0.926978417	93%

ClamAV	335	0.939748201	0.060251799	6%
QuickHeal	2567	0.538309353	0.461690647	46%
DrWEB	4554	0.180935252	0.819064748	82%
Sophos	4252	0.235251799	0.764748201	76%
eScan	5508	0.009352518	0.990647482	99%
Trend	4020	0.276978417	0.723021583	72%
Eset	4851	0.127517986	0.872482014	87%
VBA32	4851	0.127517986	0.872482014	87%

Table 4: results of antimalware from Jotti Service

From the table results we evidence that some antimalware are able to identify most of malware samples correctly (we refer to antimalware that reach a true positive percentage over the 95%), but others are very ineffective in their aim, in fact they exhibit a true positive percentage ranging from 6% to 8%.

Figure 9 shows the true percentage histograms for each antimalware.

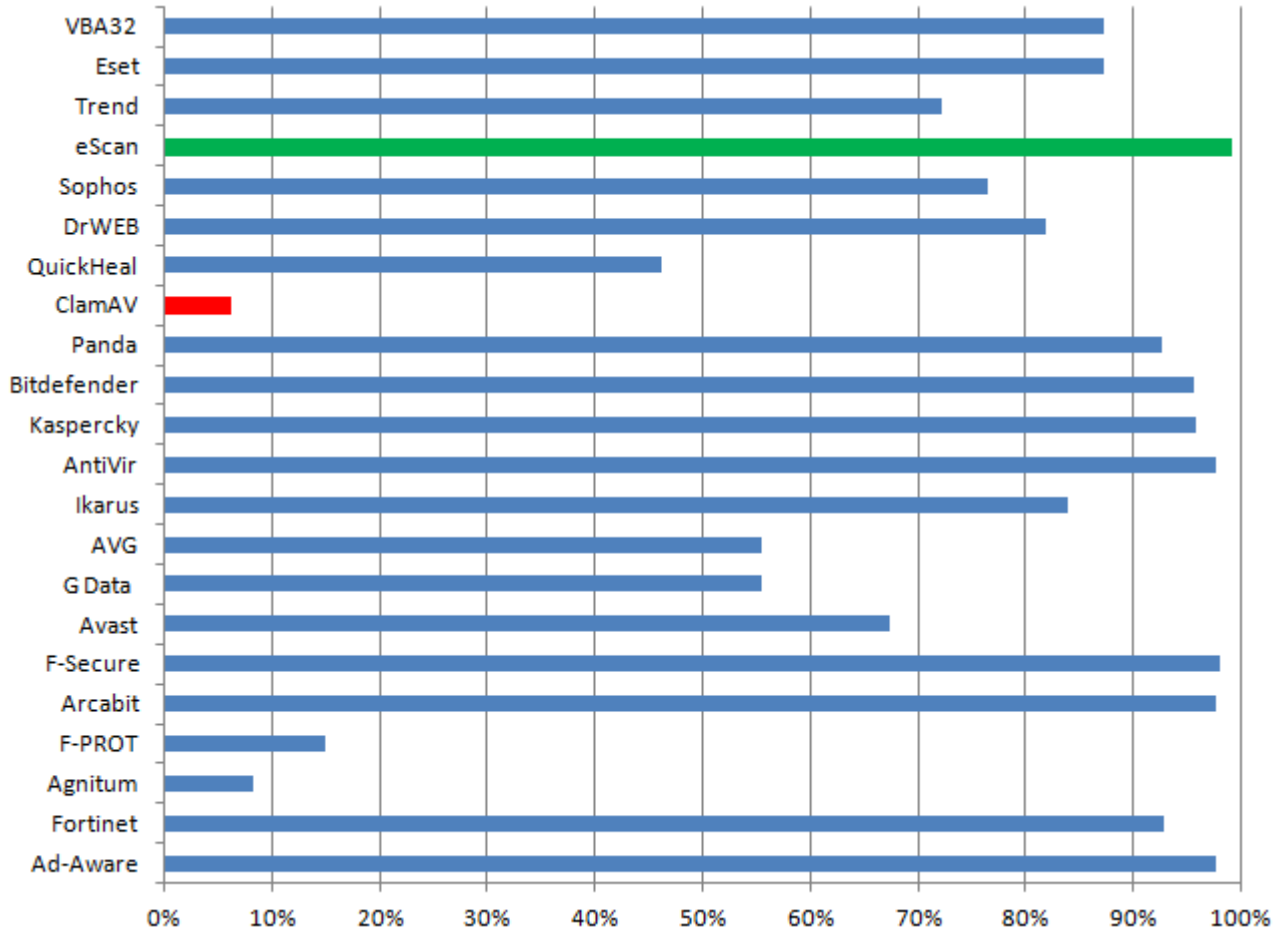

Figure 9: true positive percentage

We evidence that the antimalware with better performance is the eScan antimalware (highlighted in green) with a true positive percentage equal to 99%, while the worst one is resulted to be ClamAV (highlighted in red) with only 6% of samples correctly recognized.

The analyzed antimalware, in identifying a samples as malware, they identify also the family. Starting from previous analysis, we deduce how many elements are correctly classified in the right family. In order to make a comparison between different antimalware and then evaluate which had not only recognized the elements of the sample as malware, but also those who had properly inserted in the right family has been calculated the Euclidean distance between vectors.

We have defined the vector referred to each antimalware. This vector consists of 20 different elements, each element is one of the 20 most populous families in the dataset malware. Each vector field contains the number of elements that belong to that family.

Each vector thus obtained was compared with the vector "oracle", namely the one containing the exact number of components for each family. The Euclidean distance was calculated as follows:

$$d = \frac{\sqrt{\sum_n (o_i - a_i)^2}}{n}$$

where n represents the total number of families considered, i.e. 20; o_i is the i-th element of the oracle and a_i the i-th element of the antimalware vector.

We compute the distance for each antimalware obtaining the following result:

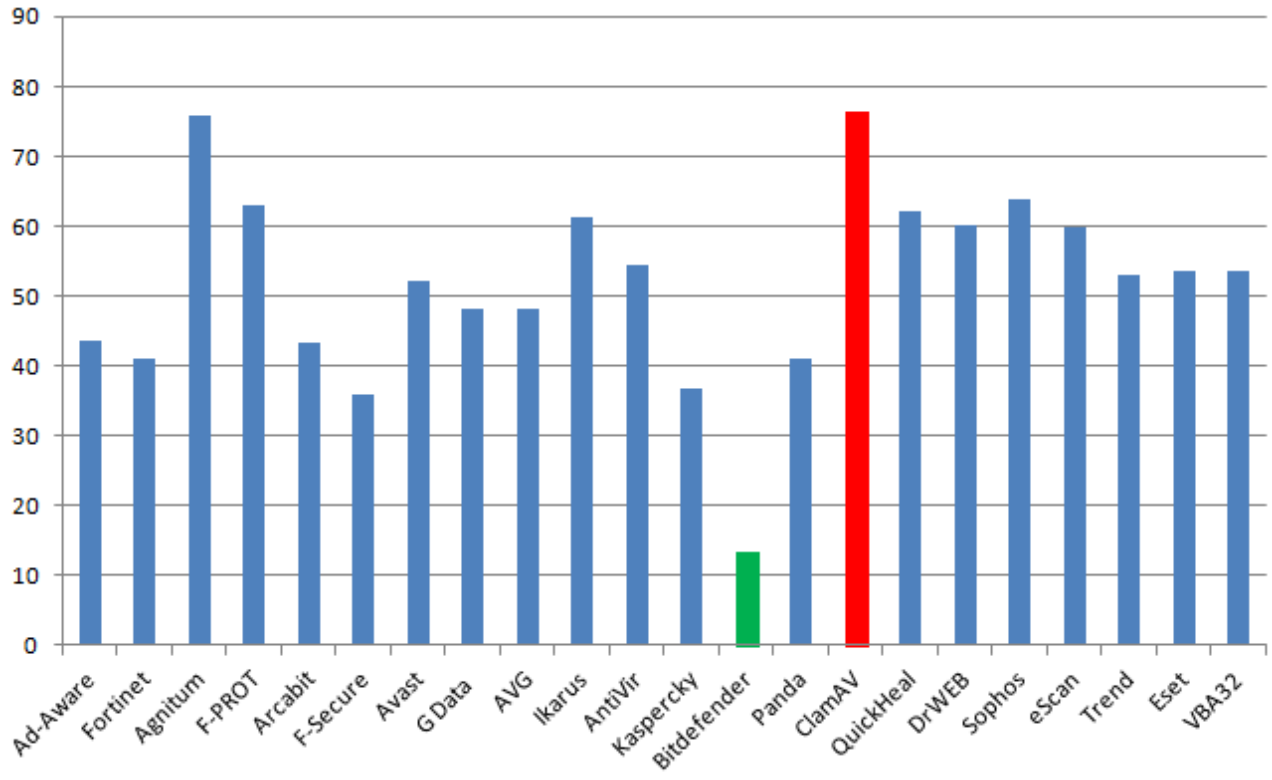

Figure 10: antimalware euclidean distance histograms

As expected, the antimalware that obtain the maximum value distance is the same that has recognized less malware samples: ClamAV (hightlighted in red). We have a trend inversion regarding the antimalware that has better classified the malware families: BitDefender (hightlighted in green). This one had obtained a percentage of true positives of 96%, then, even though it has not been the AntiMalware with a higher percentage of success, however, has been able to better classify the family samples.

We analyze also in great detail the correctness of obtained results considering for each antimalware and for each of top 20 populous families the percentages of true positives, false negatives and false positives.

True positive is the number of samples belonging to a family *x* and properly classified in that family; with false positive we consider all samples recognized as belonging to a family *y* but actually belong to another family *x*, then the set of samples correctly identified as malware but classified in the wrong family. With false negatives we consider the number of samples belonging to a family of malware, but wrongly classified as trusted.

We synthesize the results in the following histograms:

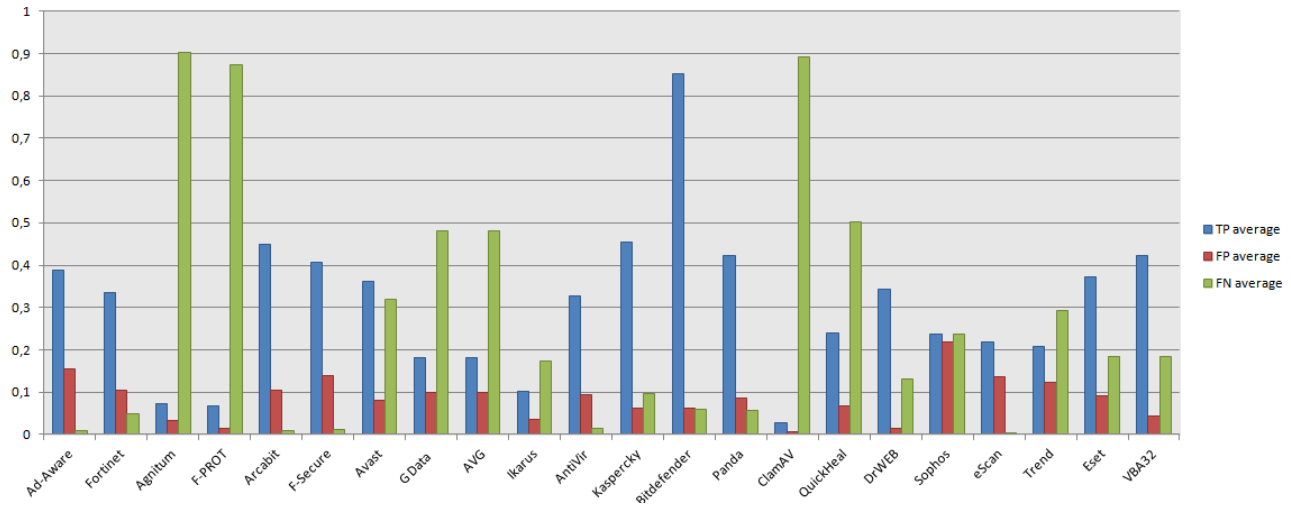

Figure 11: true positive, false positive, false negative average for antimalware histograms

Antimalware that present an high false negative average have, of course, detected the less samples number correctly.

But we must differentiate between the antimalware that correctly detected the malware and the antimalware that correctly classified the malware. In the first one the antimalware has detected the malware but with incorrect recognition of the family, while in the second case the antimalware detects also the correct family membership.

For example, eSCAN antimalware was the best in malware detection (99%) but in figure 11 we evidence that the TP and FN values are similar, this means that although it was able to identify the malware was not able, however, to classify it correctly.

Other antimalware that have a less success rate were more accurate than in the classification. For example, DrWEB antimalware while having recognized only 82% of them, however, properly classified malware samples.

In the following analysis we investigate if Jotti antimalware are able to discriminate malware, we show how many samples were detected as malware from how many antimalware.

The first column represents the number of antivirus that have marked a sample as a malware that number ranges from 0 (no anti malware detected the maliciousness of the sample) to 22 (all antimalware have classified the sample as malware), while the second one is the total number of samples that have received the malware mark.

detected by #antimalware	***#samples***
0	1
1	0
2	4
3	11
4	16
5	22
6	19
7	11
8	17
9	21
10	47
11	115
12	348
13	388
14	420
15	428
16	689
17	1078
18	833
19	740
20	265
21	69
22	18

Table 5: number of samples detected as malware with the number of antimalware that rightly have revealed this.

RQ3 response: only 7 antimalware on 22 has obtained detection percentage higher of 95%.

Evaluating the unknown dataset

In this section we consider the second dataset consists of 4000 applications downloaded from alternative stores that have been subjected to the same type of analysis performed with Andrubis, Androguard and Jotti. We describe the first results obtained considering the two stores separately and then subsequently we provide an overall picture.

Figure 12 shows the result deriving by the analysis of 2000 applications downloaded from AppChina third-party market with Andrubis.

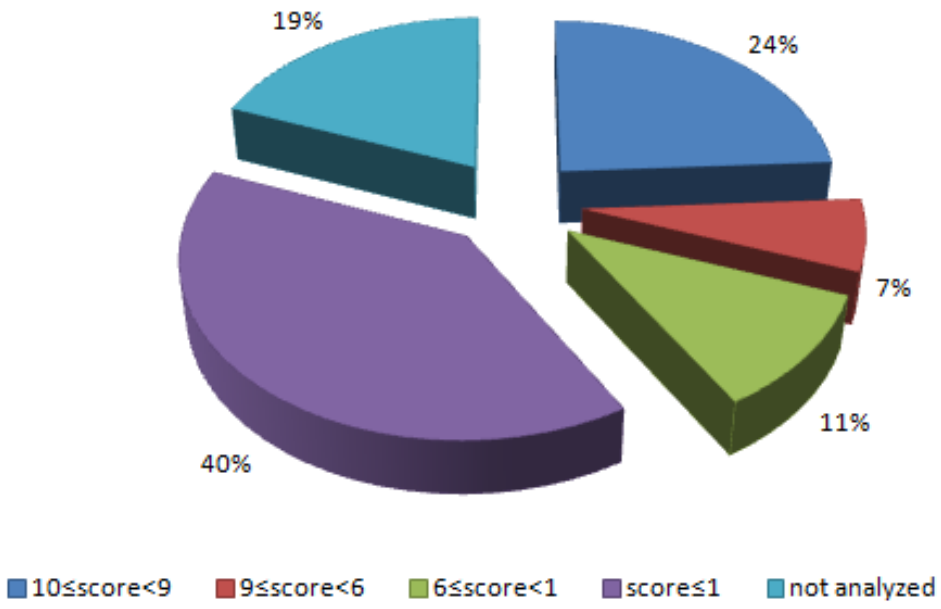

Figure 12: malware detection percentage of AppChina market using Andrubis

19% of the applications submitted were not analyzed because of the aforementioned Andrubis limitations. 51% of the samples have obtained a score of less than 6 and therefore can be classified as trusted. As much as 40% of the total received a score of less than 1, so their dangerousness is found to be particularly low. The remaining 30% obtained a score greater than 6 and then was classified as malware; in particular about 80% of the latter have received a score above 9.

It is evident that the number of detected malware is still quite relevant and downloading applications from that store could hide the pitfalls, at least according with Andrubis.

In following figure the result deriving by the analysis of 2000 applications downloaded from Gfan third-party market with Andrubis.

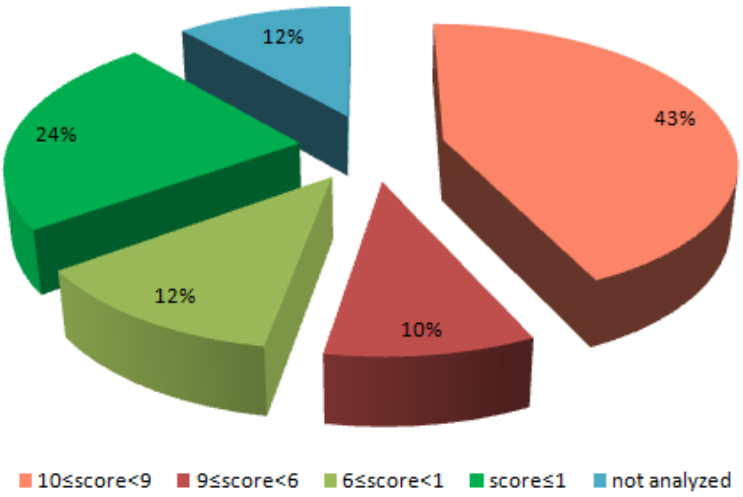

Figure 13: malware detection percentage of GFan market using Andrubis

12% of samples sent has not been achieved by the instrument; about 36% was trusted, then with a score of less than 6 and between 24% of the total has also obtained a score less than 1. As for the remaining 53% the score reported was greater than 6, and then we can consider applications classified as potentially harmful. As many as 43% of the total received a score above 9, indicating that the feedback of Andrubis was particularly severe, to emphasize the other risks which the applications may hide.

From the data analysis we notice that the percentage of malware detected for applications downloaded from Gfan is even higher than of AppChina applications.

We can now provide an overview of the two stores and have an overall picture of the analysis made by Andrubis. We added and the data obtained from Gfan and Appchina in the usual four categories of reference and we summarized combined results in the following chart:

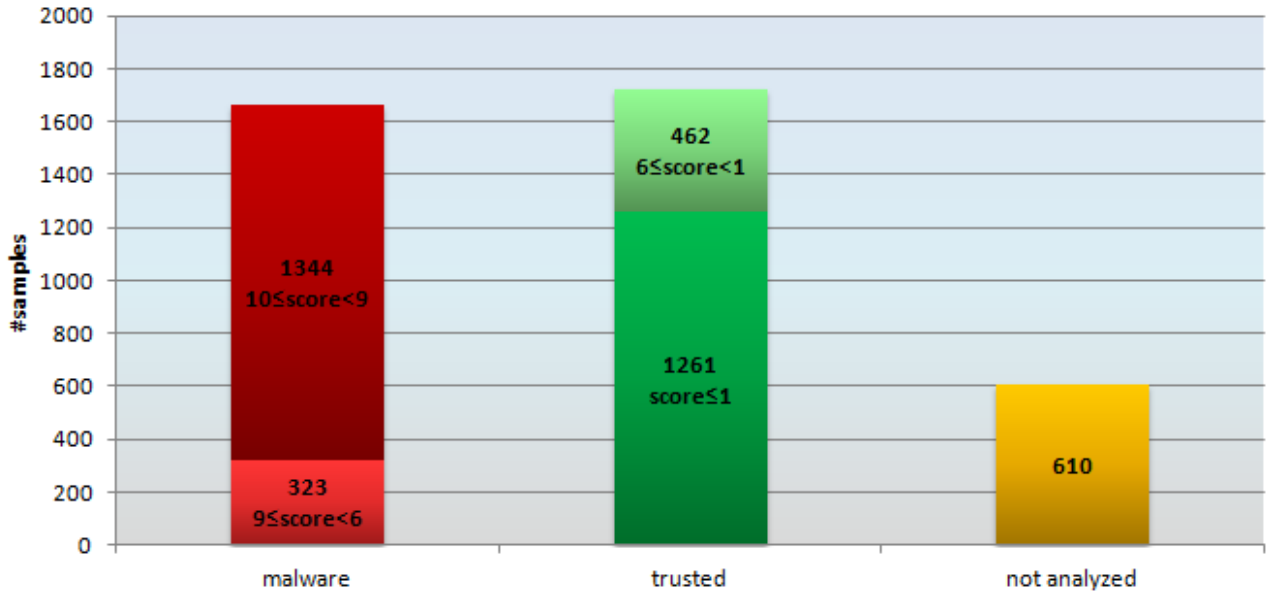

Figura 14: Histograms of the analisys on AppChina and Gfan markets with Andrubis

Excluding the 610 unused applications from Andrubis, we find that 1667 were classified as malware and 1723 as trusted. The graph also shows the different ranges of scores obtained. What is clear is that if we consider the entire set of applications downloaded from alternative stores and excludes those on which it was not possible to make an analysis, about one out of two was classified as malware, i.e. the probability of download potentially dangerous application in our devices is 50%. Obviously it is not the dataset consists of the applications downloaded by third parties labeled, it is not known a priori nature malevolent of applications, we can not verify whether the statement of Andrubis is actually correct, but at least this is what results from the data.

RQ4 response: Andrubis has classified 1667 (1344 of AppChina and 323 of Gfan) samples as malware and 1723 (1261 of AppChina and 462 of Gfan) as trusted, one out of two is a malware sample.

We also submitted the third-party dataset to Androguard, but as surprising result the tool considered all samples as trusted.

RQ5 response: Androguard has classified all samples in the third party dataset as trusted

We now analyze the results of the analysis provided by the antimalware. Regarding the 2000 applications downloaded from AppChina only 62 of them were not analyzed for the upload limitations imposed by Jotti.

Regarding the remaining 1938 applications antimalware exhibit the following results:

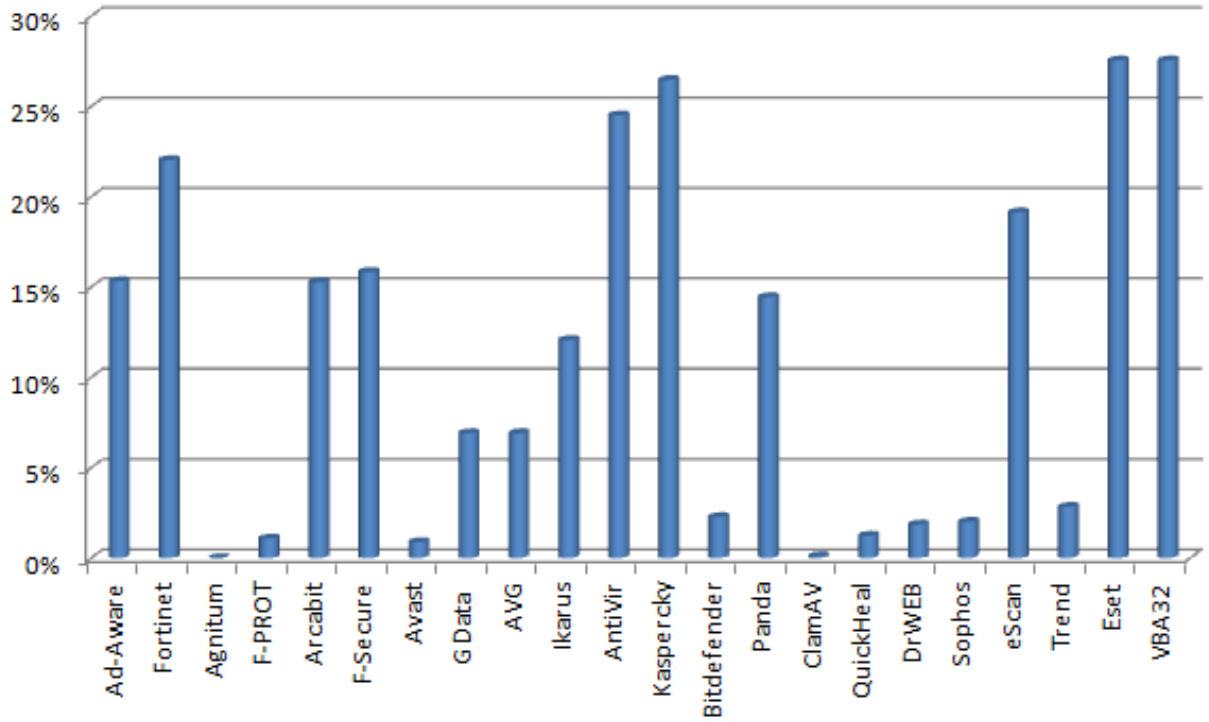

Figure 15: malware detection percentage of AppChina market using antimalwares provided by Jotty

As can be seen from Figure 15, several anti-malware have considered the entire set of applications subject as completely safe, their percentage of detection was equal to 0% or a few units higher. Other antimalware have instead found a dangerous even 25-28%.

To get an overall picture we consider in this case, as previously done with the dataset of malware, the results obtained from Jotti not in relation to individual antimalware but in relation to individual applications. Also in this case we assume that a sample detected as malicious by a number of antimalware less than or equal to three is classified as trusted, or otherwise as malware. This produces the following results:

detected by #antimalware	***#samples***
0	1247
1	61
2	62
3	39
4	60
5	71
6	54
7	65
8	20
9	25
10	55

11	67
12	52
13	47
14	11
15	1
16	0
17	1
18	0
19	0
20	0
21	0
22	0

Table 6: number of samples detected as malware with the number of antimalware that rightly have revealed this.

We synthesize the previous results in following figure:

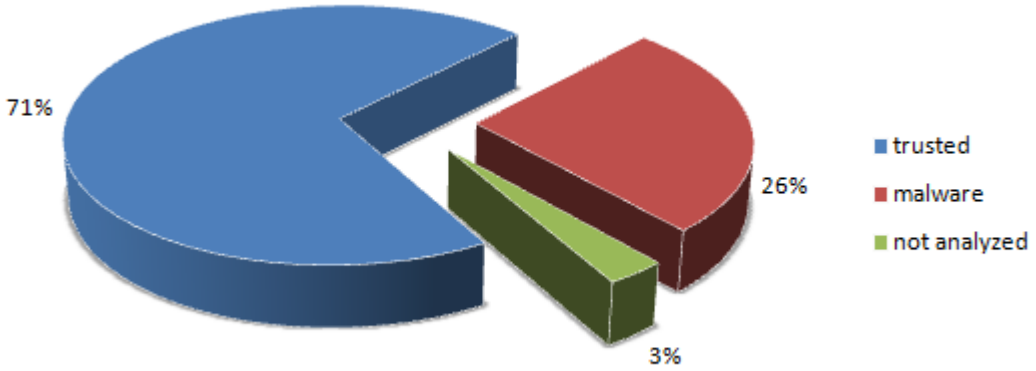

Figure 16: percentage results of AppChina market using antimalware

As explained in figure 16, 26% of the AppChina applications was discriminate as malware, while 71% is marked as trusted.

Now we investigate whether the samples detected as malware belonging to the families gathered in the dataset of malware. To find out, we calculated the number of detection for each antimalware for each family oracle. Afterwards we proceeded to add the results obtained by the different antimalware in order to have the total number of detections for families. The results are in following table:

Family	***Sum on Anti-Malware***
SMSreg	**179**
GinMaster	**82**

Plankton	**79**
Vdloader	10
Gappusin	9
Coogos	8
Glodream	4
SMSSend	4
RATC	4
Exploit.RageCage	4
DroidKungFu	1
DroidDream	1
Steek	1
Ksapp	1
DroidRooter	1
Rooter	1
Generic	1
JSmsHider	1
Anti	1
SmsSpy	1

Table 7: Families detected using Jotti antimalware on AppChina samples

We detected malware belonging to 22 of the 179 families in the dataset of known malware and especially families who have experienced a greater number of samples were *SMSreg*, *GinMaster* and *Plankton*.

We performed the analysis also on applications downloaded from Gfan third-party market, in this case only one application exceeds the upload limit and it can not be submitted.

Here the results obtained using Jotti antimalware:

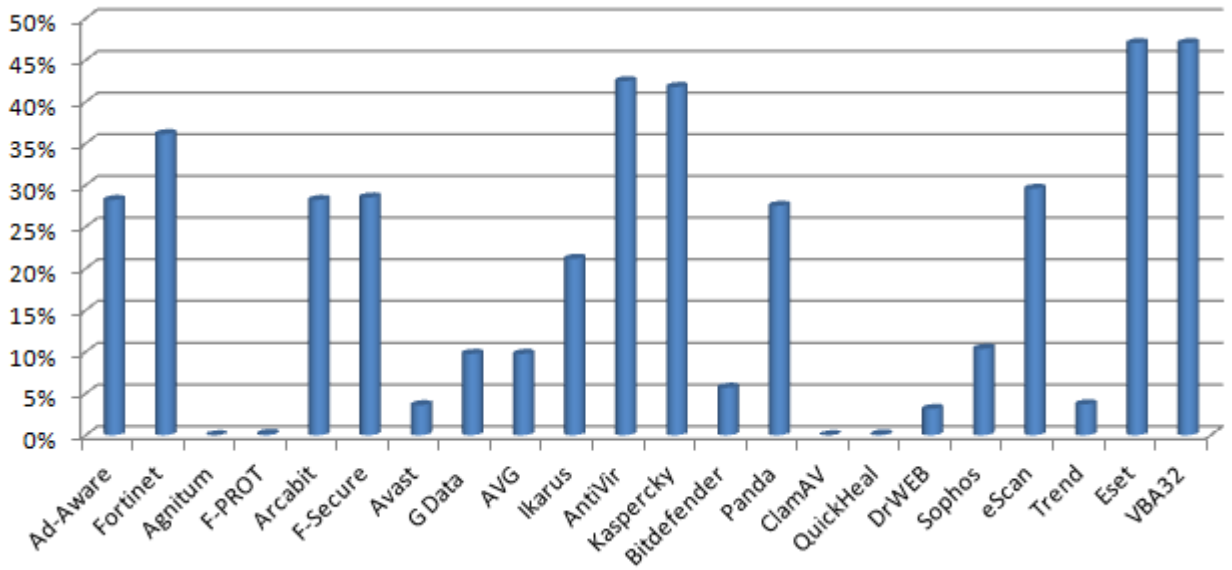

Table 8: malware detection percentage of Gfan market using antimalwares provided by Jotty

Similarly to previous case, several antimalware have obtained a percentage detection score near to 0%, these are the same antimalware that expose this result in analysing the AppChina applications. Other antimalware exhibit a detection percentage equal to 47%.

Considering the overall findings of the individual sample of our data set Jotti we achieved the following results:

detected by #antimalware	***#samples***
0	939
1	64
2	62
3	31
4	134
5	105
6	75
7	43
8	14
9	43
10	71
11	125
12	137
13	65
14	71
15	11
16	6
17	2

18	1
19	0
20	0
21	0
22	0

Table 9: number of samples detected as malware with the number of antimalware that rightly have revealed this.

Following figure summarizes the results:

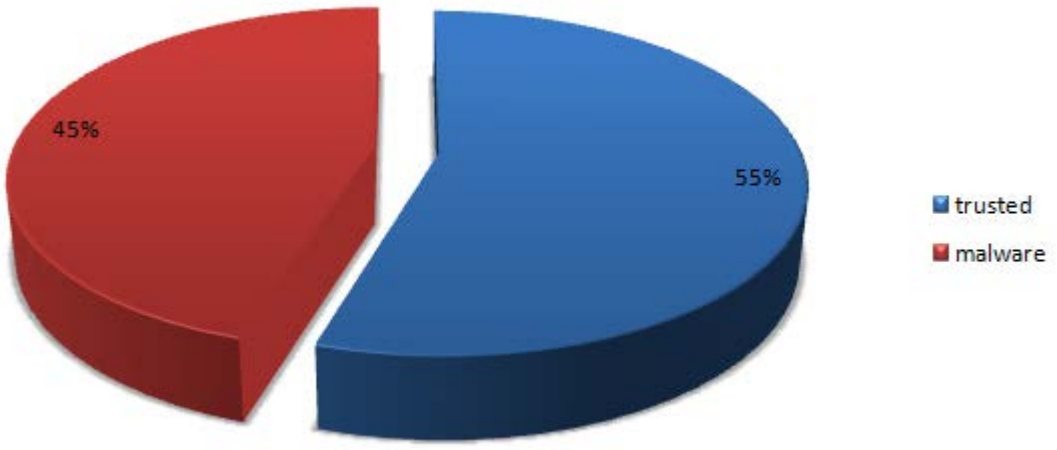

Figure 17: malware and trusted applications in the GFan applications

45% of the applications in Gfan market submitted to Jotty service has marked as malware by at least three antimalware, and thus we consider these as malicious. The other applications were marked as trusted from all antimalware.

In following table the families discovered in samples marked as malware from at least three antimalware.

Families	***Sum on Anti-Malware***
GinMaster	271
SMSreg	114
Gappusin	32
Ksapp	22
Stealer	19
Vdloader	16
SMSSend	16

Coogos	12
Fujacks	11
Glodream	9
TrojanSMS.Hippo	8
Stiniter	6
DroidRooter	5
Rooter	5
DroidKungFu	4
Nandrobox	4
BaseBridge	2
Fakengry	2
Adrd	1
RATC	1

Figura 18: Families detected using Jotti antimalware on Gfan samples

The most populous families discovered were *GinMaster* (271 samples) and *SMSreg* (114 samples).

As last analysis we compare an overview regarding the full third-party dataset (AppChina and Gfan) in order to explain the final result of Jotti analysis.

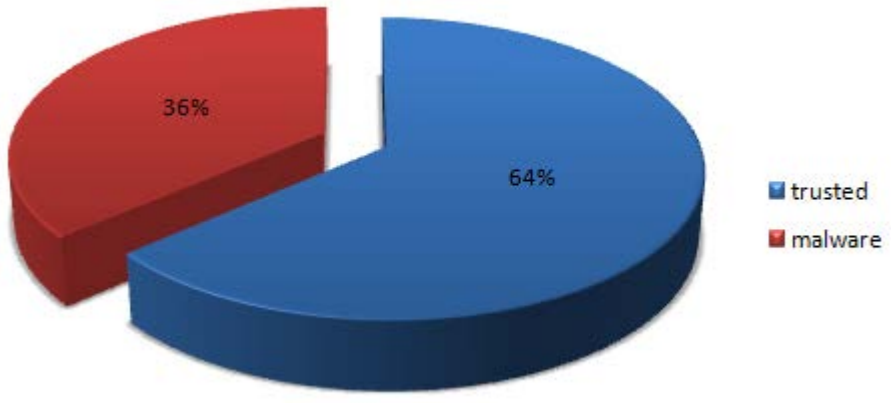

Figura 19: total results with Jotti using third-part market dataset (AppChina and Gfan)

RQ6 response: some antimalware show a percentage of malware from 25-28% (Appchina) or higher to 40% (Gfan); others antimalware evidence percentages of malware close to zero.

Conclusions

To determine the level of reliability of the current antimalware signature-based and of two of most famous research tools, we used a dataset of 5560 malware applications and others 4000 downloaded from store third-party (i.e. Appchina and Gfan).

The instruments used were Jotti, as a representative of the antimalware signature-based, which as seen, collects 22 of the most well-known antimalware currently available on the market, while Andrubis and Androguard are representative instead of prototypes of research in malware detection. Analysis of the results is seen as both Andrubis that Jotti a whole have actually been able to detect the presence of harmful software. Androguard was the worst in class.

The survey did not prove as effective as could be expected, considering the fact that the samples are dated maximum October 2012. Some Anti-Malware proved to live up to expectations obtaining detection rates of 98%, 99%, others have proved completely ineffective, even with detection rates of the order of unity. If we consider the single family membership, we can see that not everyone is able to classify them properly. Individual antimalware were not able to perform their task at best.

Regarding Andrubis, instead, the tool is actually shown in a position to make a good detection. If we neglect applications not analyzed for size or technical problems, in fact as many as 98% of the dataset was correctly identified, while Androguard exhibits a percentage equal to 22% in malware detection.

Instead considering applications from third-party store, as seen, even among these has been identified to the presence of harmful software. In particular, Andrubis found a higher percentage of malware than signature based antimalware for both store, while from Androguard point of view all the submitted third party applications were trusted samples. Obviously not knowing the true nature of the applications analyzed is not possible to evaluate the accuracy of these data, however, what emerges is that both tools have found a certain danger for these applications.

References

[1] ANDRUBIS - 1,000,000 Apps Later: A View on Current Android Malware Behaviors, https://iseclab.org/papers/andrubis_badgers14.pdf, last visit 17 April 2015

[2]Androguard, https://iseclab.org/papers/andrubis_badgers14.pdf, last visit 17 April 2015

[3]Jotti's malware scan, http://virusscan.jotti.org/it, last visit 17 April 2015

[4] Daniel Arp, Michael Spreitzenbarth, Malte Huebner, Hugo Gascon, and Konrad Rieck "Drebin: Efficient and Explainable Detection of Android Malware in Your Pocket", 21th Annual Network and Distributed System Security Symposium (NDSS), February 2014

[5] Michael Spreitzenbarth, Florian Echtler, Thomas Schreck, Felix C. Freling, Johannes Hoffmann, "MobileSandbox: Looking Deeper into Android Applications", 28th International ACM Symposium on Applied Computing (SAC), March 2013