The High-level Performance comparison of different experiments

Here, we give a comparison about different performances among these repackaged app detection systems. We compare these system from the average comparison time, coverage rate and detection rate.

The average comparison time means the time when compare two apps.

coverage rate = $\frac{the\ apps\ a\ system\ can\ hanled\ after\ removed\ the\ unnecessary\ features}{the\ whole\ apps}$

The detection rate can reflect the precision of a system.

Table I performance comparison

system	avg. comparison time	coverage rate	threshold	detection rate
DroidMOSS	0.193s	94.5%	0.8	65.3%
DroidSim	45.5s	99.8%	0.8	99.1%
ViewDroid	2.8s	69.5%		99.1%
FsQuaDRA	58ms	99.95%	0.8	27.2%
DNADroid	0.49s	57.1%	0.8	95.9%
MassVet	20.6ms	100%	0.8	88.9%

Because the ResDroid compare the similarity of two apps by change the features of apps into the sequences and use the LCS as the similarity measure algorithm. The threshold is a little different from other systems. We give the performance result of ResDroid separately.

ResDroid

Coarse Analysis result

Threshold	Number	Detection rate
0.001	14322	93.6%
0.01	15034	98.3%
0.025	15294	100%

In order to calculate the distance between the two apps, we should set the weight(the layout weight w l and the event handler weight we). Actually, according to the above description and analysis we can find that adversaries usually tend to change the Activity structure instead of the Event Handler. If w l > we , more emphasis is paid to the layout features. Otherwise, the event handler features may be regards as more important. Based on the above series of experiments, we can easily speculate that if we set w e a higher value, the false negative rate will be lower. In order to verify our speculation, we set up nine experiments, the results of the experiment as shown in TableII. According to our experimental results, we can see the first group can get the smallest false negative rate.

Table II The influence of different weight on experiment results

weight	detection rate
wl = 0.1, we =0.9	99.9%
wl =0.2, we=0.8	99.8%
wl =0.3, we=0.7	99.3%
wl =0.4, we=0.6	98.5%
wl =0.5, we=0.5	97.6%
wl =0.6, we=0.4	97.0%
wl =0.7, we=0.3	96.3%
wl =0.8, we=0.2	96.0%
wl =0.9, we=0.1	95.7%

Here, we also give readers the detailed performance results of each systems.

DroidMOSS Detection result

coverage rate 94.5%   14446/15291

Similar value	no.	Detection rate
80	9986	65.3%
75	10341	67.63%
70	11013	72.0%

DroidSim

coverage rate
15260/15291 = 99.8%

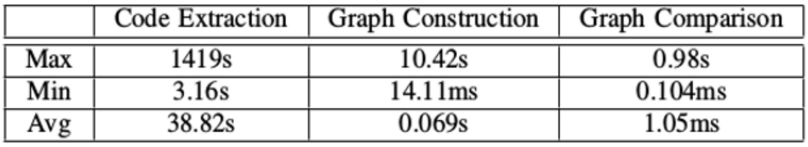
Similar value	no.	Detection rate
1	15015	94.3%
0.9	15134	98.9%
0.8	15163	99.1%

ViewDroid

comparative time 2.80s

coverage rate 10635/15191 69.5%
detection rate = 10544/10635 = 99.1%

Table III Execution time of ViewDroid exclude nine abnormal apps



	Code Extraction	Graph Construction	Graph Comparison
Max	1419s	10.42s	0.98s
Min	3.16s	14.11ms	0.104ms
Avg	38.82s	0.069s	1.05ms

FsquaDRA

comparative time 184.32s
avg comparative time 58 ms

process time 2942.73s

coverage rate 15284/15291 = 99.95%

FSquaDRA processed all the repackaged app pairs just cost 2942.73s. It cost 184.32s to compare all the repackaged app pairs. The average comparing time is 58ms, the longest comparing time is 244ms and the shortest comparing time is 9ms. It takes 1215.52s to get the apk features to the memory.

Threshold	Number	Detection rate
0.9	3351	23.2%
0.8	4158	27.2%
0.7	4857	30.0%
0.6	5627	36.9%

DNADroid

coverage rate 57.1%

Threshold	Number	Detection rate
0.9	7989	91.5%
0.8	8378	95.9%
0.7	8531	97.7%

MassVet

coverage rate 100% 15295

Threshold	Number	Detection rate

0.9	12696	83%
0.8	13599	88.9%
0.7	13909	90.9%