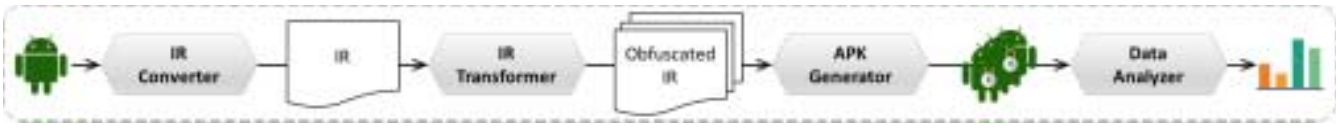

Figure 1: Obfuscation study methodology

Table 2: Obfuscation strategies of each obfuscation tool

	Trivial				Non-trivial						
Obfuscator/Strategy	ALIGN	DR	MAN	REPACK	CF	CR	ENC	IDR	JUNK	MR	REF
Apktool/Jarsigner		✓		✓							
Allatori					✓		✓	✓		✓	
DashO					✓		✓	✓			
DroidChameleon			✓		✓	✓	✓		✓		✓
ADAM	✓				✓		✓	✓	✓		
ProGuard								✓			

- Allatori [3] is a commercial Java and Android obfuscation tool that supports a wide range of obfuscation strategies. Many companies such as Amazon, Fujitsu, and Motorola rely on Allatori to protect their software systems from being reverse engineered. The providers of this tool, Smardec Inc., provided us with a full version for educational purposes.
- ProGuard [22] is a widely used open-source shrinker, optimizer, obfuscator, and preverifier for Java bytecode. A preverifier performs certain checks on Java bytecode prior to runtime. ProGuard supports identifier renaming and is the default tool in many development environments, including Android Studio [14], the official IDE for Android apps.
- ADAM [47] is a research tool for obfuscating Android apps. It transforms the Smali code of a reversed-engineered app.
- DroidChameleon [44] is a state-of-the-art research tool for obfuscating Android apps which supports a wide range of obfuscation strategies. Compared to ADAM, DroidChameleon supports more complex transformations. Like ADAM, DroidChameleon transforms the Smali code of a reversed-engineered app.
- DashO [15] is a commercial tool for obfuscating Android and Java applications. DashO provides static analysis protection and runtime security control against tampering, unauthorized debugging, and some runtime attack patterns. This tool supports control-flow, string-encryption, and identifier-renaming transformations. The providers of DashO, PreEmptive Solutions, supplied us with a full free version valid for 30 days.
- Apktool and Jarsigner were used to perform the DR and REPACK obfuscation strategies, respectively. These two transformations often work in tandem because a reassembled APK must be resigned.

We also considered another tool, DexGuard [17], which is an advanced and commercial version of ProGuard. We contacted the providers of DexGuard to obtain an educational or commercial version of their tool to run on our dataset. Unfortunately, they only allow their tool to run on a restricted number of Android apps; and they do not sell licenses for research purposes. Hence, we did not include it in this study.

4.3 Evaluation Framework

To conduct our study, we have developed the framework depicted in Figure 1, which consists of the following four modules: *IR Converter*, *IR Transformer*, *APK Generator*, and *Data Analyzer*. *IR Converter* takes an Android APK as input and converts its code to Intermediate Representation (IR) formats. *IR Transformer* utilizes all obfuscation tools to transform the IR format using a variety of obfuscation strategies. *APK Generator* repackages each obfuscated IR file and generates an obfuscated APK from that file. *Data Analyzer* scans obfuscated apps using anti-malware products, stores the scanning results in a MySQL database, analyzes the scanning results, and creates various statistical reports.

Our framework is reusable and extendable. A user can add new obfuscation tools and support different obfuscation strategies. Therefore, we make the framework available for researchers and practitioners [21]. The framework is a Python program that consists of more than 5,500 lines of code, not counting the obfuscation tools.

IR Converter. Obfuscation tools do not require source code and they work directly on the IR format. Therefore, this module converts an APK file to two IR formats: *smali* using Baksmali and Java bytecode using *dex2jar*. In our framework, we generate these two IR formats since ADAM and DroidChameleon work on *smali* code while all other obfuscation tools work on Java bytecode.

IR Transformer. This module generates several obfuscated IR files of the original IR file. The framework is configured to leverage twenty nine different obfuscation strategies using seven obfuscation tools (recall Section 4.2).

APK Generator. For each obfuscated IR file, this module generates an obfuscated Android app. First, this module leverages the *dx* tool from the Android SDK to convert an obfuscated IR to a *classes.dex* file. Next, it generates an APK file with the new *classes.dex* using Apktool. Finally, the APK file is signed using *jarsigner* with our own certificate, since the original certificate of the app cannot be obtained.

Data Analyzer. This module uses the VirusTotal service to scan apps using anti-malware products. This module uploads the apps to VirusTotal, which scans them using more than 60 up-to-date commercial anti-malware products. For each uploaded app, VirusTotal returns a unique scanning ID, which Data Analyzer uses later to download the scanning reports and stores them in a MySQL database. Data Analyzer queries and processes the database to generate various statistical reports.

4.4 Anti-malware Products

We have evaluated the accuracy and the resiliency of 61 commercial anti-malware products against obfuscations. Due to space limitations and to ensure readability, we focus on the results of the 21 most popular Android anti-malware products in this paper; however, we make the results for all 61 anti-malware products available online [21].

Table 3 shows the anti-malware products evaluated in this study and includes the following information for each product: its number of *Downloads*; its overall user satisfaction score as represented using a star-rating (*Stars*); and the number of users who reviewed the product (*Reviewers*). The numbers in Table 3 are obtained from Google Play.

5 DATA ANALYSIS AND RESULTS

For conducting our experiments, we have leveraged a high performance computing cluster (HPC), managed by our organization, that has more than 200 compute nodes with a total of more than 8,000 cores. Each compute node has 264GB-512GB RAM. We utilized HPC

Table 3: Anti-malware products (K: Thousand. M: Million)

Product	Downloads	Stars	Reviews	Product	Downloads	Stars	Reviews
Ikarus	100K - 500K	4.2	2,862	Trustlook	10M - 50M	4.4	476,671
Emsisoft	100K - 500K	4.2	1,425	McAfee	10M - 50M	4.4	506,491
Fortinet	100K - 500K	4.2	2,086	Avira	10M - 50M	4.5	441,016
AegisLab	100K - 500K	4.2	2,905	Norton Symantec	10M - 50M	4.5	946,230
F-Secure	500K - 1M	4.1	12,183				
Comodo	500K - 1M	4.6	33,395	ESET-NOD32	10M - 50M	4.7	490,840
GData	1M - 5M	4.0	8,850	Kaspersky	10M - 50M	4.7	2,061,983
Sophos	1M - 5M	4.3	11,816	DrWeb	50M - 100M	4.5	1,044,410
TrendMicro	1M - 5M	4.6	49,977	Antiy-AVL	100M - 500M	4.1	2,166
BitDefender	5M - 10M	4.5	88,809	Avast	100M - 500M	4.5	4,724,478
CAT-QuickHeal	5M - 10M	4.4	204,709	AVG	100M - 500M	4.5	5,785,171

to run thousands of jobs simultaneously. On each app, we applied 29 different obfuscation strategies: 4 trivial transformations, 7 non-trivial transformations, and 18 combined transformations. Table 4 shows the number of obfuscated apps resulting from applying the 29 obfuscation strategies leveraged by the obfuscation tools. An empty cell indicates an obfuscation strategy that is not support by a particular obfuscation tool. In total, we have generated 73,362 obfuscated apps from 3,000 benign apps and 3,000 malicious apps.

In the remainder of this section, we present the results of our experiments. We measured the effectiveness of anti-malware products at identifying malicious apps in terms of their *precision*, which measures the extent to which benign apps are labeled as malicious, and *recall*, which measures the extent to which malicious apps are labeled as benign. We use the *F-score*, i.e., the harmonic mean of precision and recall, to measure the overall detection rate of anti-malware products.

5.1 RQ1. Obfuscation Strategies

We studied the accuracy of anti-malware products with respect to a wide variety obfuscation strategies in two scenarios. In the first scenario (Section 5.1.1), we compare the detection rates of each anti-malware product on the original dataset and the obfuscated dataset. In the second scenario (Section 5.1.2), we measure the detection rate of anti-malware products against each obfuscation strategy.

5.1.1 Detection rate on original and obfuscated apps. Figure 2 shows the detection rate of 21 anti-malware products on the original dataset of 6,000 apps, depicted as black bars, and the obfuscated dataset of 73,362 apps, depicted as gray bars. Figure 2 demonstrates that the detection rate of anti-malware products on the original dataset is above 85% for 16 products, and between 75% and 85% for 4 anti-malware products. *TrendMicro* exhibits the lowest detection rate, 56%. The average detection rate is 87% on the original dataset. Consequently, prior to application of obfuscation strategies from obfuscation tools, these top anti-malware products are quite effective at protecting Android users; albeit there is room for improvement.

Once obfuscation strategies are applied, the detection rates for those anti-malware products decrease significantly, as shown in Figure 2. For example, *AegisLab* achieves the highest detection rate on the original dataset, 96%, since it mislabeled only 247 apps in the original dataset. Its detection rate has dropped to 55% on the obfuscated dataset—a 40% decrease—as it mislabeled 27,636 apps. Other anti-malware products are also severely impacted by code obfuscation. While the average detection rate of anti-malware products on the original apps is 87%, the average detection rate on the obfuscated dataset is 67%—a 20% decrease.

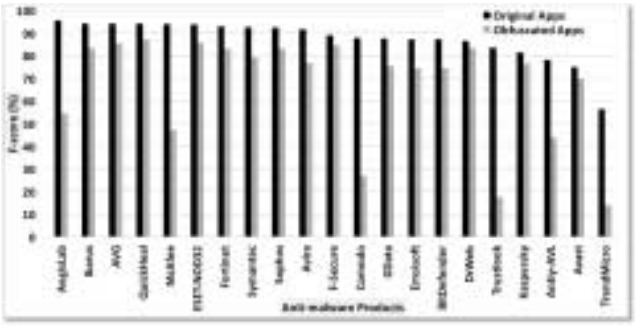

Figure 2: Detection rate of 21 anti-malware products on 6,000 original apps and 73,362 obfuscated apps.

Finding 1: Code obfuscation significantly impacts Android anti-malware products. The average detection rate for the top anti-malware products decreases from 87% to 67%—a 20% decrease.

5.1.2 Detection rate for each obfuscation strategy. To better understand the impact of every obfuscation strategy on each anti-malware product, Table 5 presents the detection rate of anti-malware products, expressed as the F-score, on the original dataset and the various obfuscation strategies. For example, the detection rate of *Symantec* on the original dataset is 93%. This detection rate has dropped to 64% on obfuscated apps using MAN, to 69% on obfuscated apps using ENC, and to 31% on obfuscated apps using ENC_IDR.

Table 5 demonstrates that the majority of anti-malware products are not affected by REF. We consider a transformation's effect on a product's detection rate to be negligible if the detection rate has either improved, decreased by less than 3%, or remains above 85%. In fact, the accuracy of *F-Secure*, *GData*, *BitDefender*, and *Emsisoft* improves on apps obfuscated using REF. This result indicates that the intensive use of *code reflection* makes an app look suspicious to our studied anti-malware products, improving their detection rates. Unfortunately, this phenomenon may result in false positives for certain anti-malware products. For instance, *AVG* erroneously marked 307 benign apps obfuscated using REF as malicious, while also correctly detecting nearly all malicious apps obfuscated using REF.

Finding 2: REF transformations make apps look suspicious, increasing the chance of an app being labeled as malicious.

Perhaps most surprising is that certain trivial obfuscation strategies are quite effective against the top anti-malware products. Notably, the anti-malware products that we studied rely heavily on analyzing an app's manifest file, which contains configuration information. Consequently, these products are often evaded by apps obfuscated using MAN, which involves the trivial addition or modification of permissions or Intent filters. For example, the detection rate of *McAfee* dropped from 94% to 21% for apps obfuscated using MAN. Overall, the average detection rate of anti-malware products fell to 60% from 87% when apps are obfuscated using MAN—a 28% decrease.

Finding 3: MAN, which is a trivial obfuscation strategy, severely impacts many anti-malware products, on average, decreasing a product's detection rate by 28%.

Another interesting, possibly counter-intuitive, conclusion that we can draw from Table 5 is that combined transformations are not always superior to individual transformations. For instance, while

Table 4: Number of obfuscated apps using the obfuscation strategy in the column leveraged by the obfuscator in the row.

	Trivial				Non-trivial							Combined Strategies																			
Obfuscator/Strategy	ALIGN	DR	MAN	REPACK	CF	CR	ENC	IDR	JUNK	MR	REF	CF_ENC	CF_IDR	CF_MAN	CF_MR	CR_MAN	ENC_IDR	ENC_MAN	JUNK_MAN	MAN_REF	CF_CR_MAN	CF_ENC_IDR	CF_ENC_MR	CF_IDR_MR	CF_REF_MAN	ENC_REF_MAN	CF_ENC_IDR_MR	CF_CR_ENC_MAN_REF	CF_CR_ENC_JUNK_MAN_REF	Total apps
Apktool/Jarsigner		3,693		5,880																										9,573
Allatori					1,612		1,613	1,609		1,612					1,609								1,607	1,607			1,606			12,875
DashO					1,094		1,089	1,097				1,082	1,084				1,077					1,083									7,606
DroidChameleon			3,597		2,593	1,952	687		351		1,487	610		1,594		1,752		679	349	1,385	1,361					993	658		570	314	20,932
ADAM	5,487				0	4,175	2,708	4,119	4,182																						20,671
ProGuard								1,705																							1,705
Total apps	5,487	3,693	3,597	5,880	5,299	6,127	6,097	8,530	4,533	1,612	1,487	1,692	1,084	1,594	1,609	1,752	1,077	679	349	1,385	1,361	1,083	1,607	1,607	993	658	1,606	570	314	73,362

Table 5: (RQ1) Detection rate of anti-malware products, measured by their F-score (%), against each obfuscation strategy.

		Trivial				Non-trivial							Combined Strategies																	
Anti-malware	Original	ALIGN	DR	MAN	REPACK	CF	CR	ENC	IDR	JUNK	MR	REF	CF_ENC	CF_IDR	CF_MAN	CF_MR	CR_MAN	ENC_IDR	ENC_MAN	JUNK_MAN	MAN_REF	CF_CR_MAN	CF_ENC_IDR	CF_ENC_MR	CF_IDR_MR	CF_REF_MAN	ENC_REF_MAN	CF_ENC_IDR_MR	CF_CR_ENC_MAN_REF	CF_CR_ENC_JUNK_MAN_REF
AegisLab	96	84	52	36	92	52	35	53	39	62	66	58	41	11	57	66	37	3	68	61	58	43	4	65	66	66	68	63	58	35
Ikarus	94	95	94	66	96	75	84	81	86	86	93	88	75	77	84	88	85	64	89	86	89	86	60	84	87	93	89	82	92	87
CAT-QuickHeal	94	95	93	92	94	89	**91**	75	88	89	93	94	73	**92**	93	93	91	55	91	90	94	91	54	76	91	94	84	**70**	89	80
AVG	94	75	96	63	96	79	83	85	91	84	97	88	77	**77**	83	97	85	58	90	85	89	85	57	92	96	92	90	**92**	91	85
McAfee	94	90	50	21	93	47	20	45	52	16	73	23	20	41	17	73	20	21	22	20	22	17	22	66	66	15	18	63	14	20
ESET-NOD32	94	94	92	91	94	93	93	80	91	46	94	66	75	92	92	94	93	68	86	46	66	91	61	80	92	68	59	72	80	57
Fortinet	93	94	89	86	93	88	83	78	84	75	91	86	67	79	91	88	83	58	87	77	87	84	50	83	83	89	75	72	72	56
Symantec	93	86	87	64	88	76	84	69	79	84	92	88	63	68	83	92	85	31	90	85	89	85	31	75	90	92	90	73	91	85
Sophos	93	93	91	90	93	89	85	70	79	93	88	92	70	84	93	87	86	52	91	95	92	87	50	66	72	93	91	51	91	85
Avira	92	92	87	84	92	85	84	65	78	78	87	60	61	78	86	86	85	38	80	80	59	83	38	69	85	58	33	63	73	61
F-Secure	89	87	87	85	90	85	82	81	84	95	90	94	73	65	91	90	82	53	93	94	92	84	53	87	88	93	93	84	80	71
Comodo	88	88	27	16	82	22	17	19	24	17	19	15	17	20	11	33	16	20	14	18	16	11	20	20	26	9	14	23	11	18
GData	88	91	84	75	88	79	61	75	77	95	85	91	62	54	79	85	50	46	83	79	81	50	45	79	77	81	82	77	57	41
BitDefender	87	90	84	73	88	78	60	74	75	95	85	91	61	46	76	85	45	44	83	78	78	46	43	79	77	77	83	77	58	41
Emsisoft	87	90	84	73	88	78	60	74	75	95	85	91	61	46	76	85	45	43	83	78	78	46	43	79	77	77	83	77	59	41
DrWeb	87	88	83	81	88	89	86	90	86	93	88	40	92	89	90	88	86	90	94	94	41	90	89	88	87	39	40	87	36	35
Trustlook	84	10	23	0	48	17	0	22	20	0	36	0	2	3	0	38	0	1	0	0	0	0	2	40	39	0	0	40	0	0
Kaspersky	81	83	75	70	82	81	76	70	75	88	81	80	73	77	81	81	77	50	86	89	81	81	50	70	77	83	84	64	91	85
Antiy-AVL	78	79	56	26	80	41	22	47	47	13	68	18	20	25	12	70	21	24	12	8	20	12	25	70	69	12	15	65	8	7
Avast	75	75	63	57	75	73	66	66	66	78	74	75	71	67	78	74	68	46	91	79	76	76	45	60	69	83	91	47	91	86
TrendMicro	56	57	11	7	48	12	7	10	14	6	16	10	9	15	6	16	7	11	5	7	10	5	10	12	14	7	4	12	3	5
AVERAGE	**87**	**83**	**72**	**60**	**85**	**68**	**61**	**63**	**67**	**66**	**76**	**64**	**55**	**57**	**66**	**77**	**59**	**42**	**68**	**64**	**63**	**60**	**41**	**68**	**73**	**63**	**61**	**64**	**59**	**51**

the detection rate of *AVG* against CF is 79%, its detection rate against combined transformations that include CF is between 57% and 97%.

Finding 4: In general, combined transformations do not affect detection rates more than single transformations: The average detection rate of anti-malware products is 61% for single non-trivial obfuscations, and 61% for combined obfuscations.

Figure 3 contains box-and-whisker plots illustrating the impact of each obfuscation strategy on all anti-malware products. These results suggest that some obfuscation strategies have negligible effects on the majority of anti-malware products. For example, REPACK did not affect 19 anti-malware products. Similarly, the use of the MR transformations did not affect 14 anti-malware products. Lastly, the REF transformation did not thwart the majority of anti-malware products.

Figure 3 demonstrates that ENC_IDR and CF_ENC_IDR are very effective in thwarting anti-malware products. In fact, these two transformations evaded all anti-malware products except *DrWeb*.

Finding 5: ENC_IDR and CF_ENC_IDR are the most successful transformations for evading anti-malware products.

5.2 RQ2. Obfuscation Tools

For RQ2, we studied the detection rate of anti-malware products on apps transformed using various obfuscation tools. To that end, we analyze the results of each anti-malware product's detection rate on each obfuscation tool. We further assess the overall effect of each obfuscation tool across all studied anti-malware products.

Table 6 depicts the detection rate of each anti-malware product on apps transformed using each obfuscation tool. From Table 6, we