Android Malware Source Code Analysis

Ramón Costales
Juan Tapiador

Motivation & Objectives

Is Android malware production more complex and becoming an industry?

Manually collect and analyze samples to obtain:

- Dataset insights
- Code size
- Code quality
- Development costs
- Comparison

Dataset

Malware Types, Tags, No. Samples, Permissions, Capabilities, VirusTotal Detections

Acquisition & Analysis

Acquisition

#01. Github Searches

#02. Underground Forums

#03. Malware Databases: vx-underground, theZoo, sppen...

#04. Following web search links

Manual Analysis

97 SAMPLES
3,538,683 SLOCS

Malware Types

Malware Type	No. Malware Samples
RAT	31
Spyware	13
Trojan-Spy	9
Keylogger	8
Trojan-Banker	7
Rootkit	5
Locker	4
Ransomware	4
Phishing	3

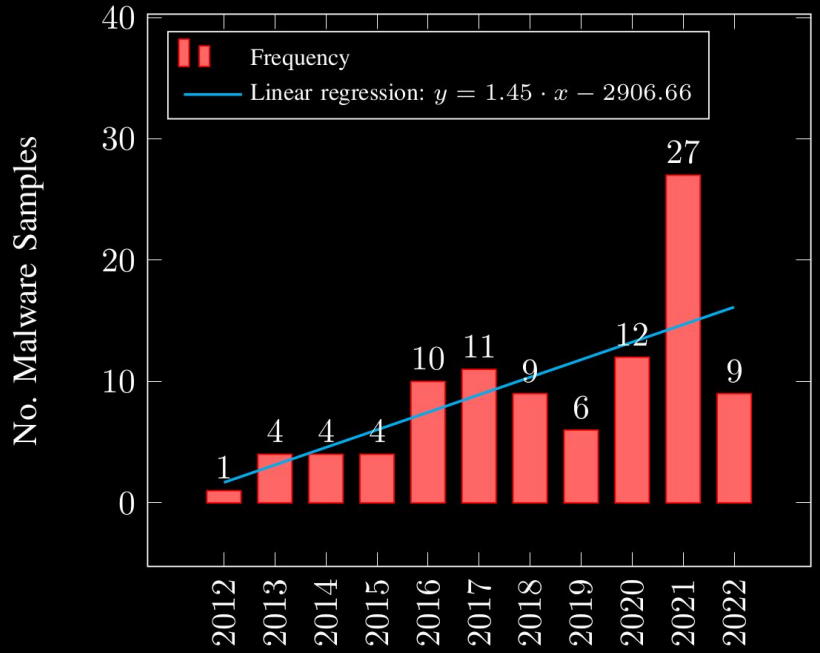
Malware Type	No. Malware Samples
Trojan-SMS	3
Dropper	2
Trojan-Backdoor	2
Backdoor	1
Downloader	1
Password-Stealing-Ware	1
Scareware	1
Trojan	1
Trojan-Wiper	1

Malware Tags

Malware Tag	No. Malware Samples
Spyware	72
Botnet	60
Backdoor	44
C2	44
Billing-Fraud	40
Trojan	35
RAT	34
Downloader	31
Elevated-Privilege-Abuse	27

Malware Tag	No. Malware Samples
Locker	19
Keylogger	17
Mailfinder	12
Wiper	12
Password-Stealing-Ware	11
Phishing	9
Encryption-Ransomware	8
Screen-Locking-Ransomware	8
Overlay	7

Samples By Year



Malware Permissions

Permission	No. Malware Samples
INTERNET	66
RECEIVE BOOT COMPLETED	55
WRITE EXTERNAL STORAGE	51
READ SMS	47
ACCESS NETWORK STATE	45
READ PHONE STATE	44
READ CONTACTS	44
SEND SMS	39
WAKE LOCK	35

Permission	No. Malware Samples
ACCESS FINE LOCATION	35
RECEIVE SMS	33
RECORD AUDIO	31
READ EXTERNAL STORAGE	31
CAMERA	31
CALL PHONE	26
READ CALL LOG	24
ACCESS COARSE LOCATION	21
SYSTEM ALERT WINDOW	20

Malware Capabilities

01

Steal Information

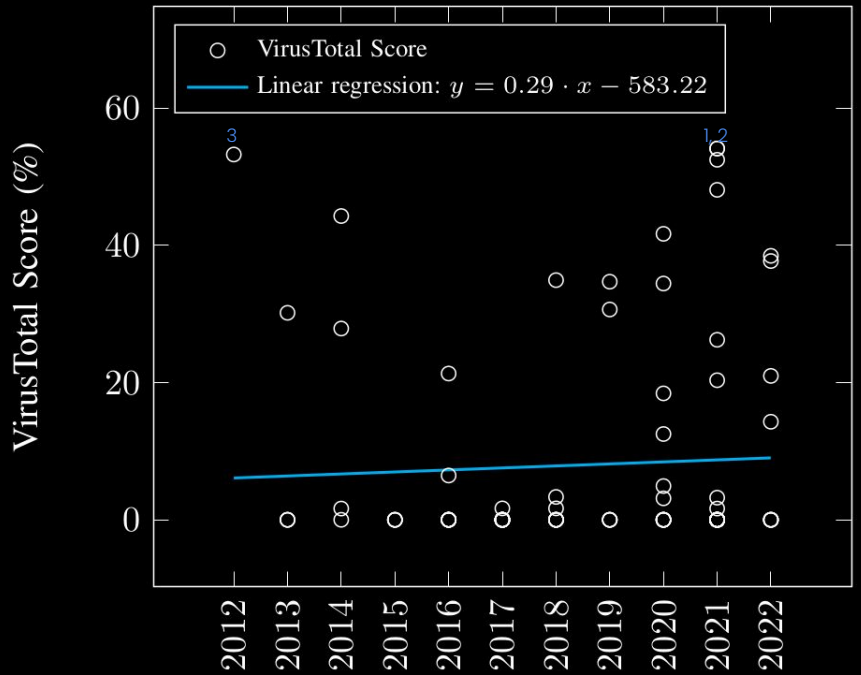
- Upload and List Files
- List Installed Apps
- Get Tasks
- Input Capture
- Screenshot
- Read SMS
- Read Contacts
- Camera

02

Control the Device

- Download and Delete Files
- Install, Uninstall and Open Apps
- Encrypt and Decrypt Files
- Lock the Device
- Hide the App Icon
- Remote Shell
- Draw Over Other Apps
- Make Phone Calls

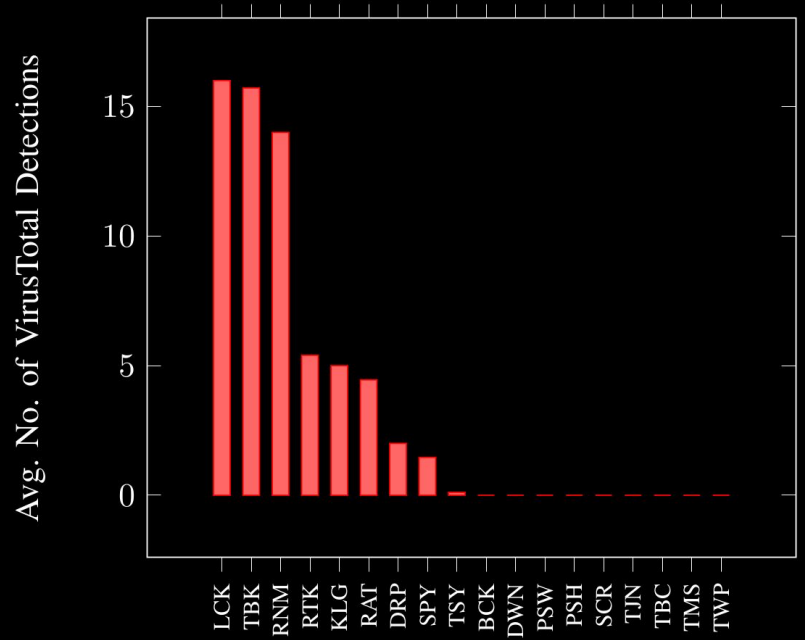
VirusTotal Detections By Year



1	Locker	andr0id_l0cker.MX	54.1%
2	Trojan-Banker	Cerberus.d	54.1%
3	RAT	AndroRAT	53.2%

AVERAGE: 8.026%

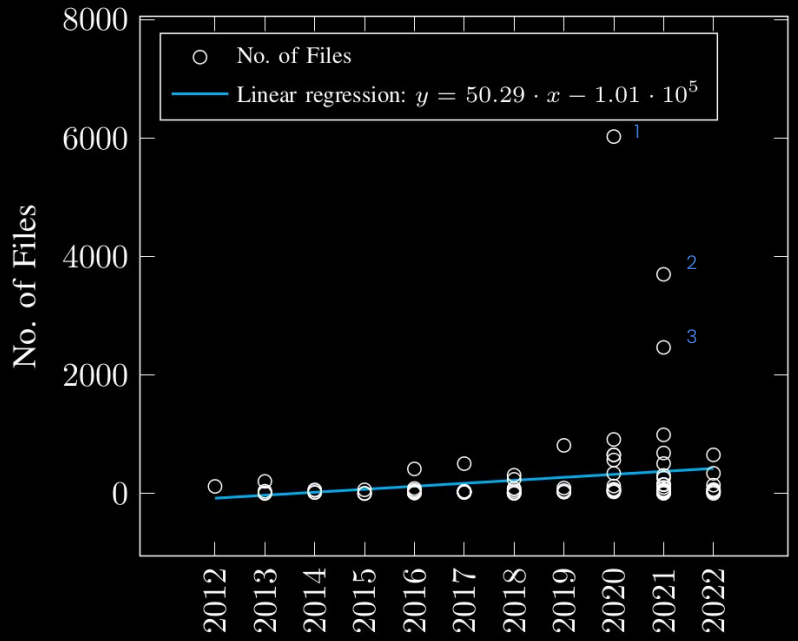
VirusTotal Detections By Type

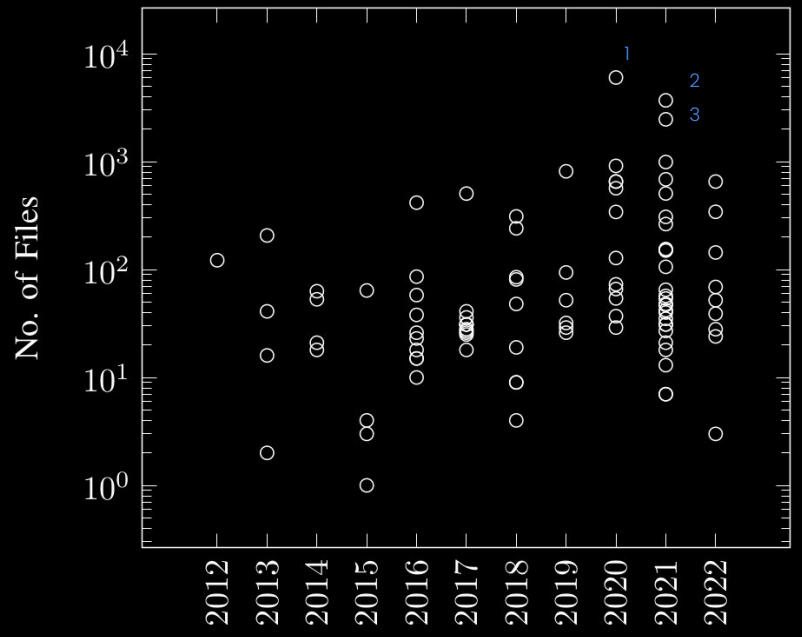


Code Size

No. Files, SLOCs, No. Functions, Languages

Files By Year

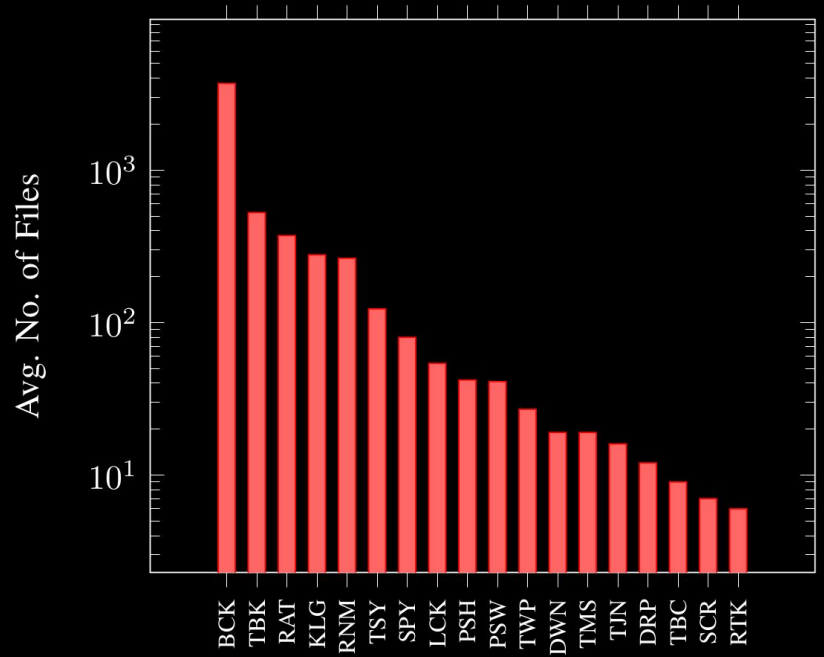




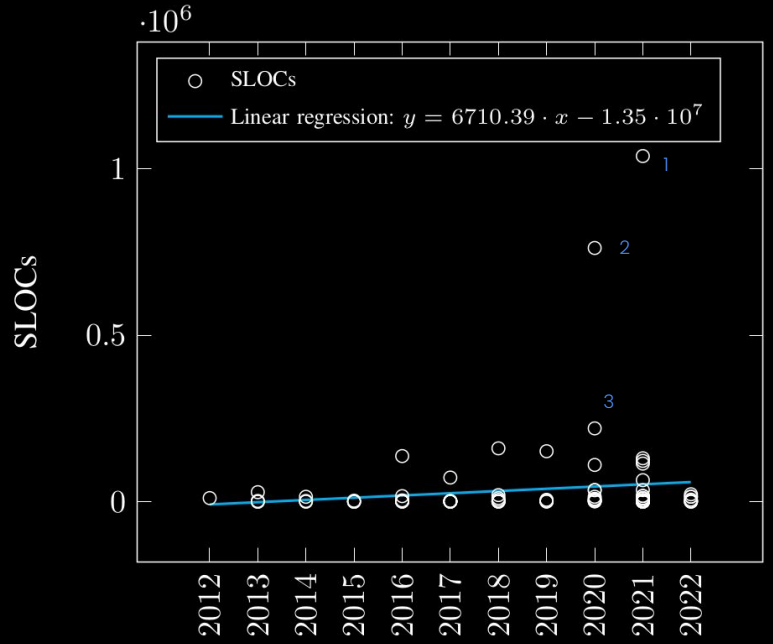
AVERAGE: 256.74

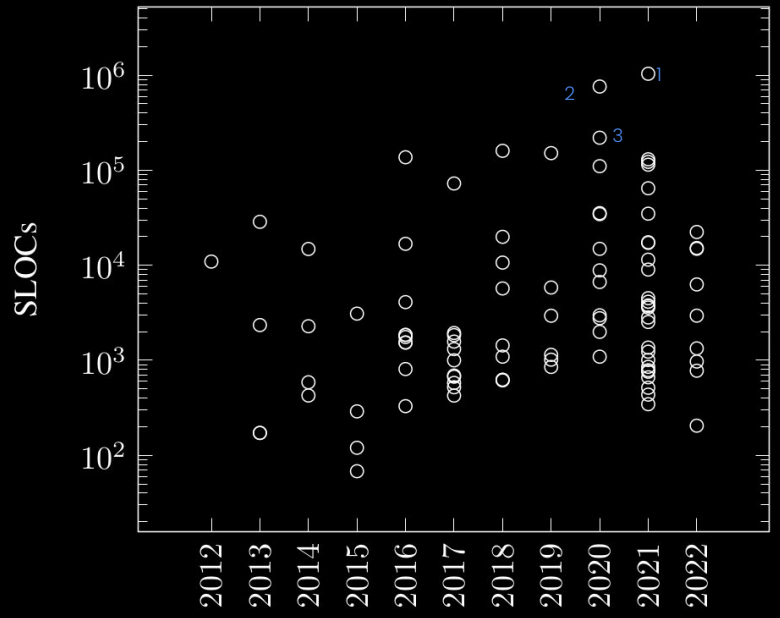
1	RAT	AhMyth	6,024
2	Backdoor	Bootloader-Backdoor	3,700
3	RAT	Arbitrium	2,467

Files By Type



SLOCs By Year

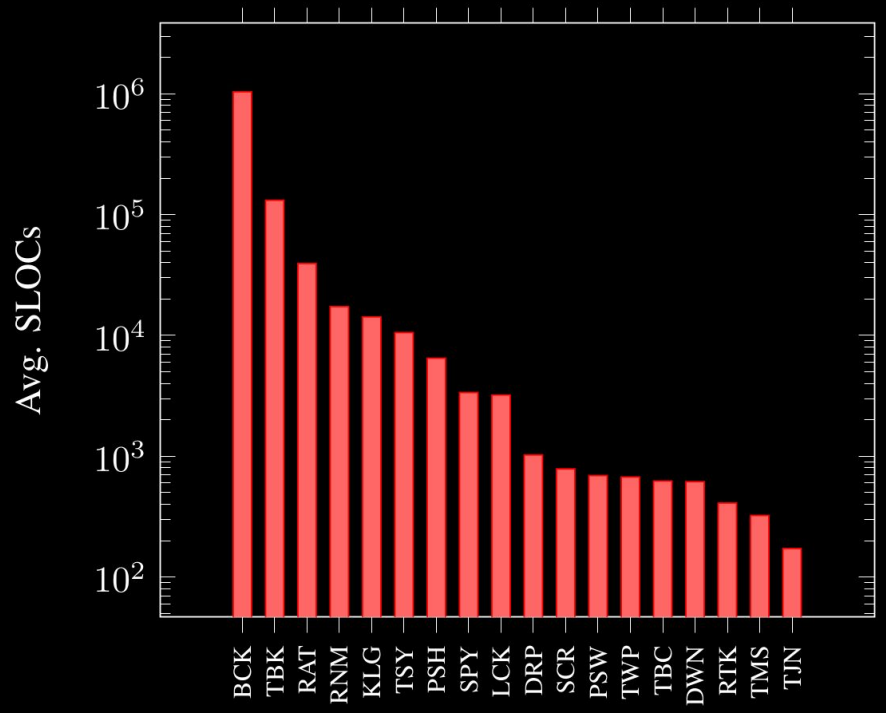




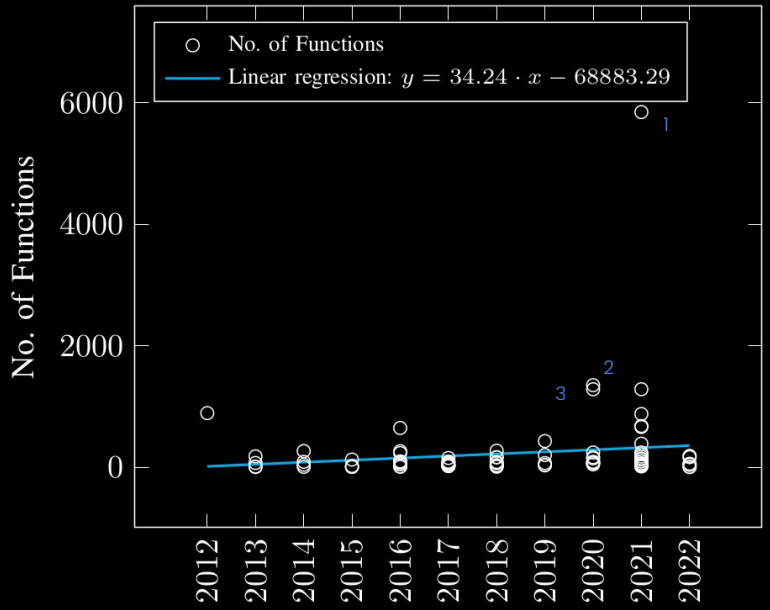
AVERAGE: 36,481.27

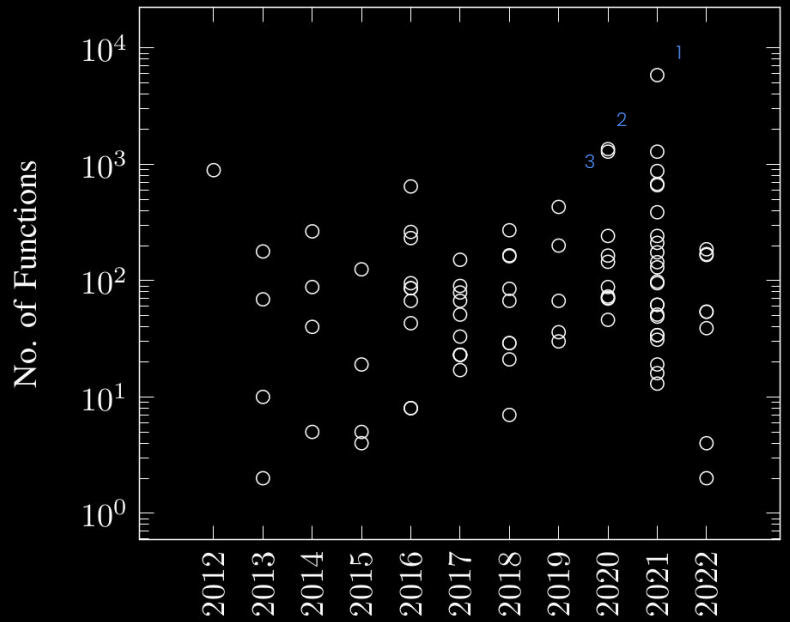
1	Backdoor	Bootloader-Backdoor	1,037,561
2	RAT	AhMyth	761,610
3	Trojan-Banker	Cerberus	220,173

SLOCs By Type



Functions By Year

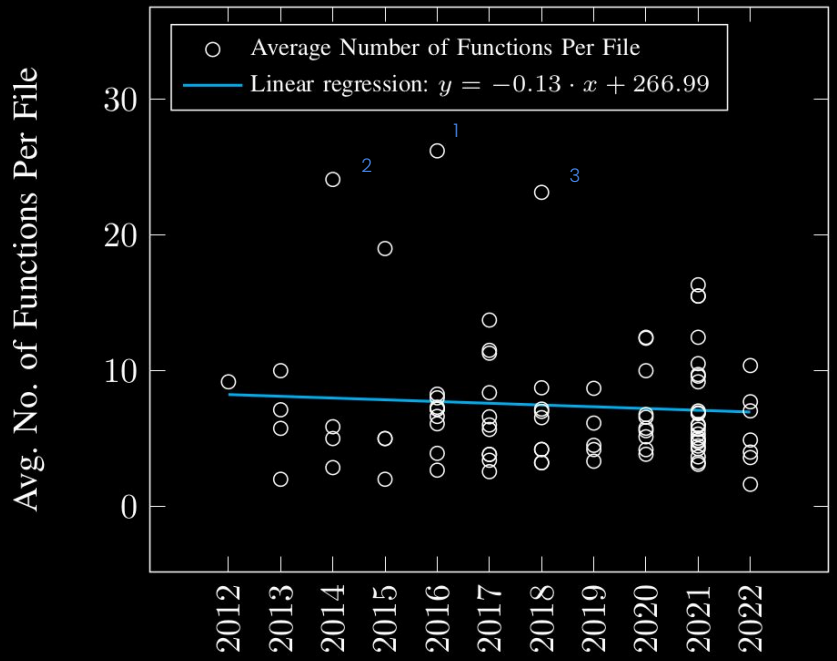




AVERAGE: 233.85

1	Ransomware	Covid-Locker	5,846
2	Keylogger	Lokiboard-mod	1,350
3	Keylogger	Lokiboard	1,283

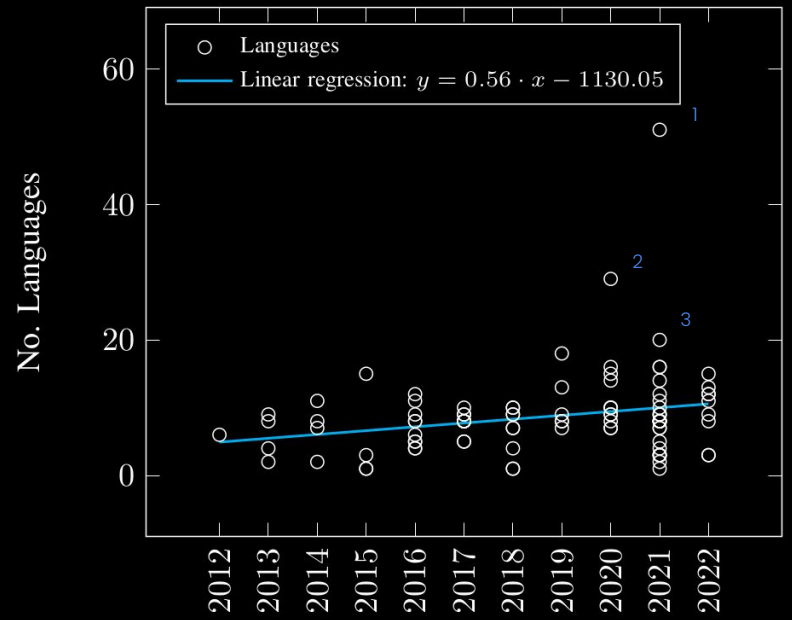
Avg. Functions Per File



1	RAT	BetterAndroRAT	26.20
2	RAT	Dendroid	24.09
3	RAT	rdroid	23.14

AVERAGE: 7.33

Programming Languages



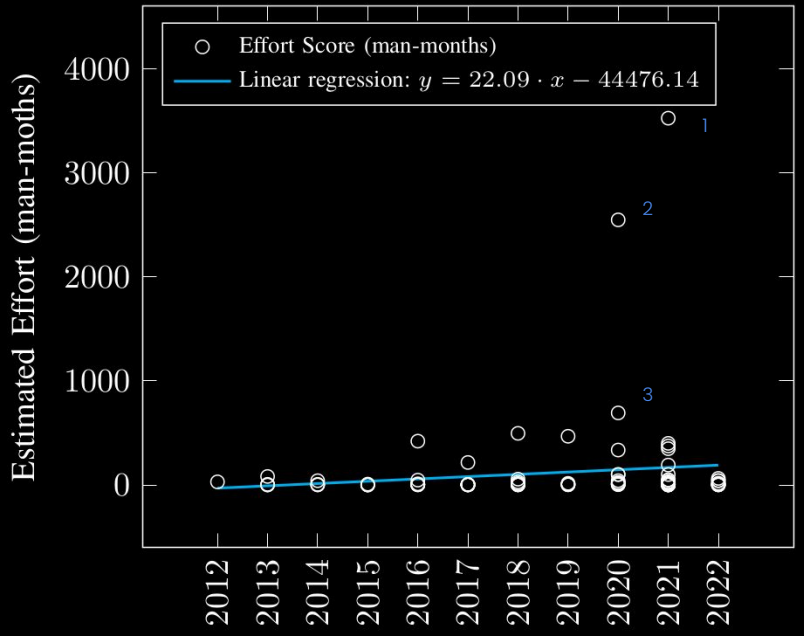
1	Backdoor	Bootloader-Backdoor	51
2	RAT	AhMyth	29
3	RAT	Arbitrium	20

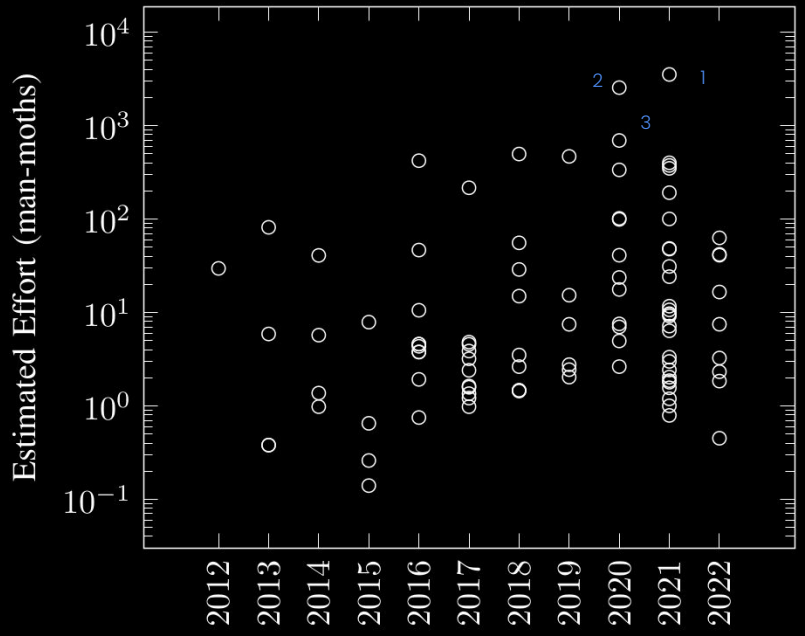
AVERAGE: 8.75

Development Costs

Effort, Development Time, Team Size

Effort

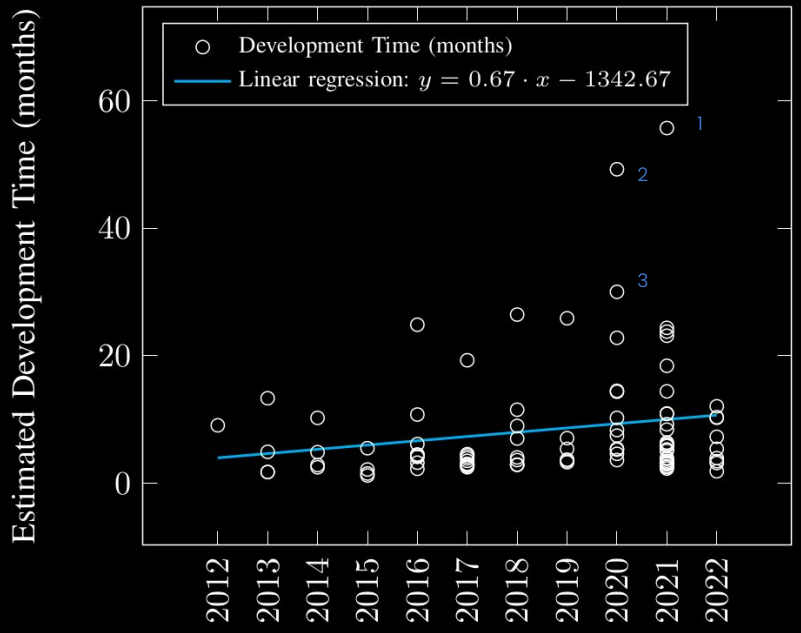


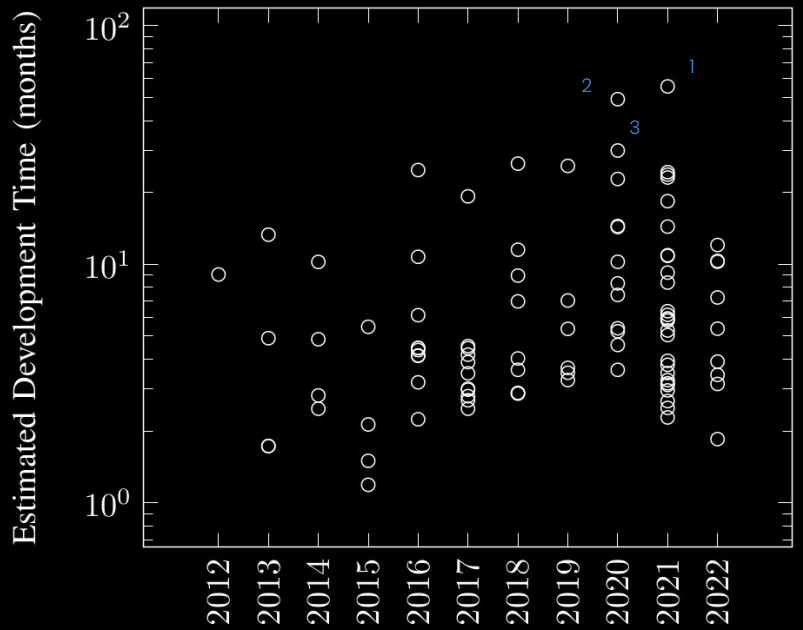


AVERAGE: 115.98

1	Backdoor	Bootloader-Backdoor	3,523.92
2	RAT	AhMyth	2,547.01
3	Trojan-Banker	Cerberus	692.01

Development Time

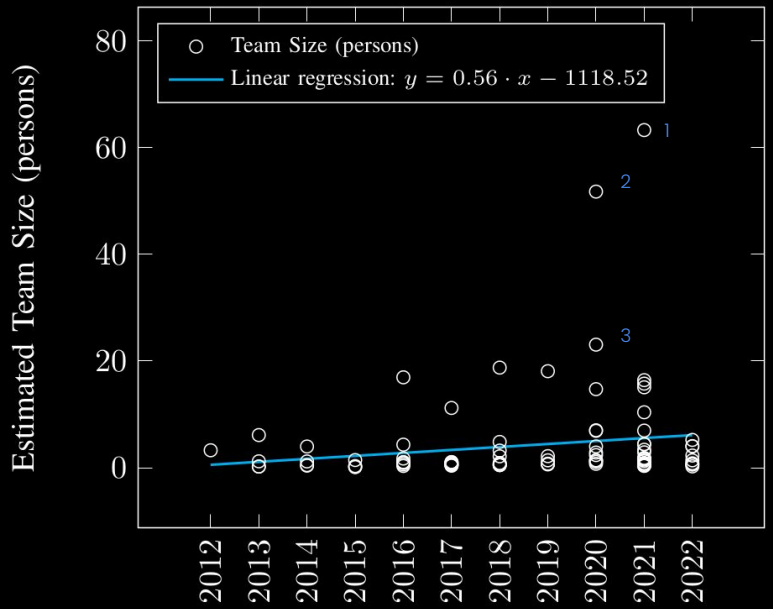


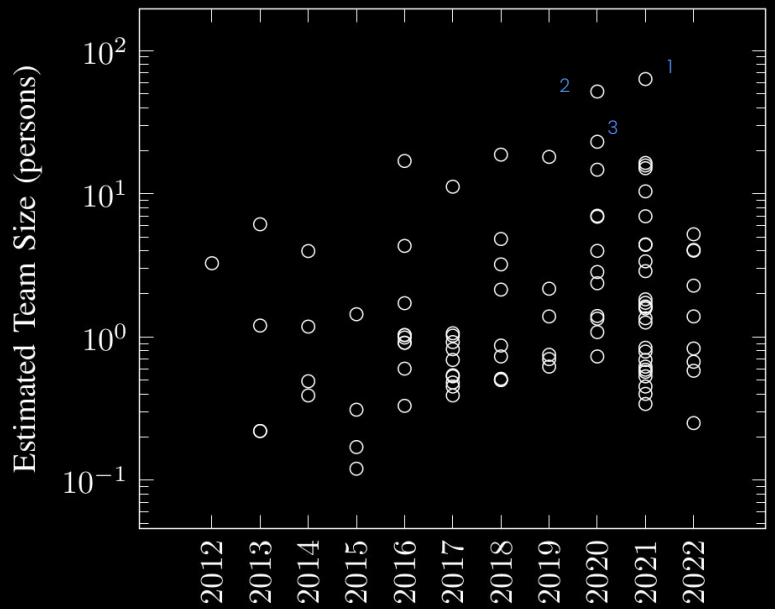


AVERAGE: 8.32

1	Backdoor	Bootloader-Backdoor	55.69
2	RAT	AhMyth	49.23
3	Trojan-Banker	Cerberus	30

Team Size





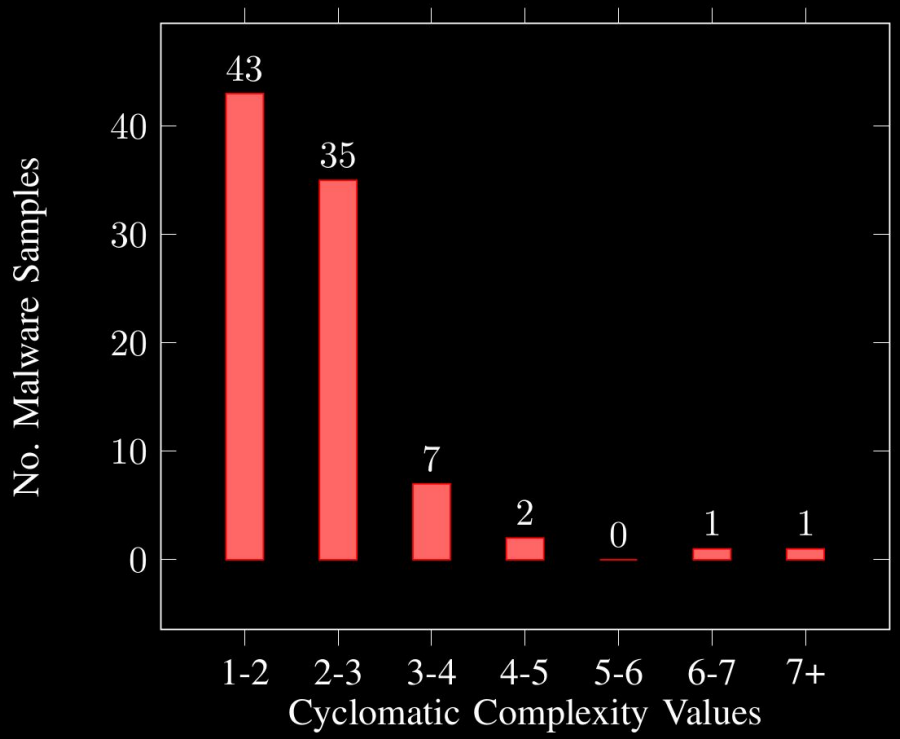
AVERAGE: 4.31

1	Backdoor	Bootloader-Backdoor	63.27
2	RAT	AhMyth	51.74
3	Trojan-Banker	Cerberus	23.06

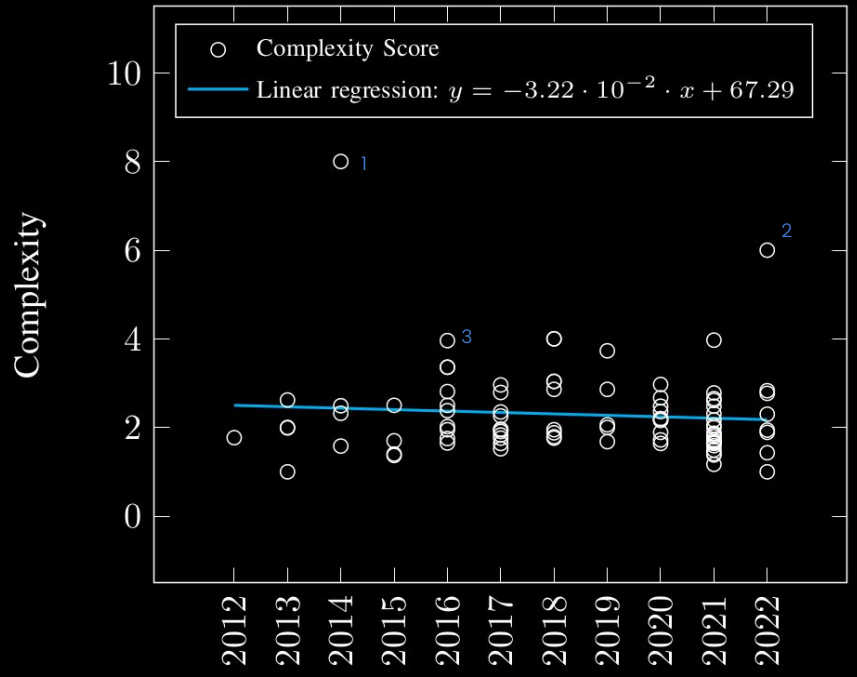
Code Quality

Complexity, Maintainability, Density of Comments

Complexity By Values



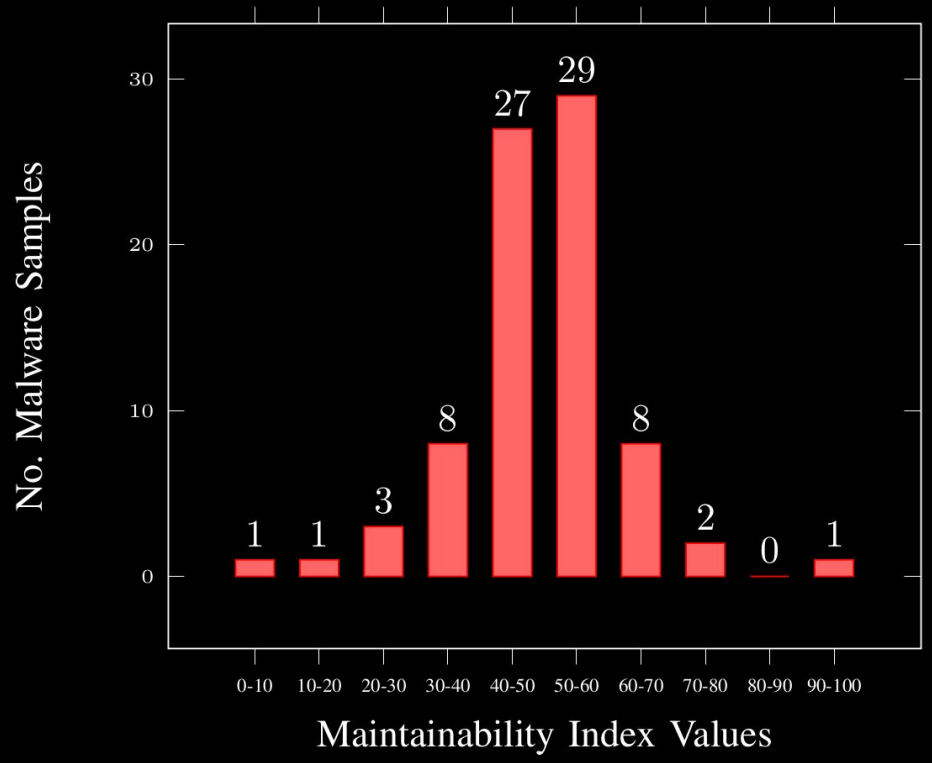
Complexity By Year



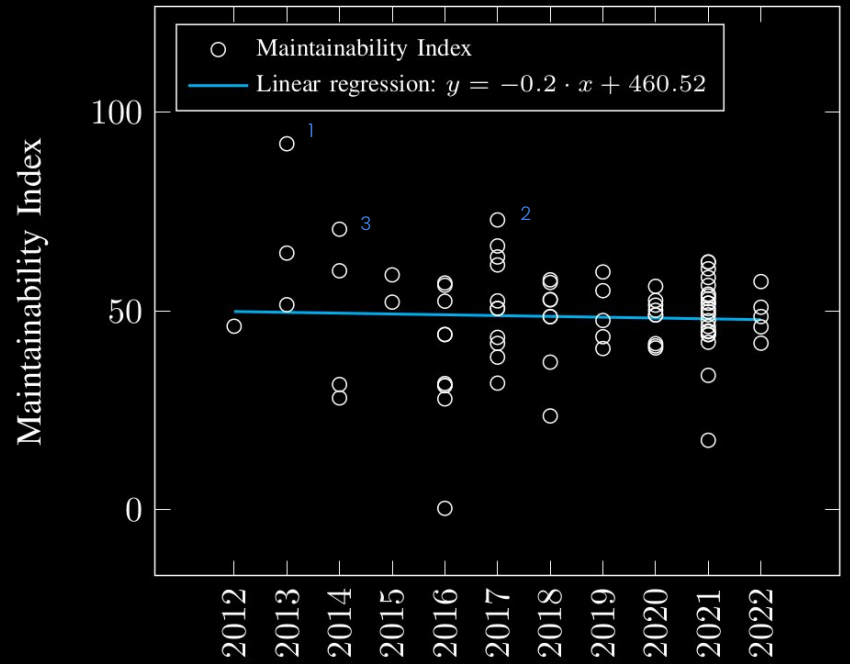
1	Rootkit	Adore	8
2	Ransomware	SARA	6
3	Trojan-Backdoor	DarkSilent	4

AVERAGE: 2.29

Maintainability By Values



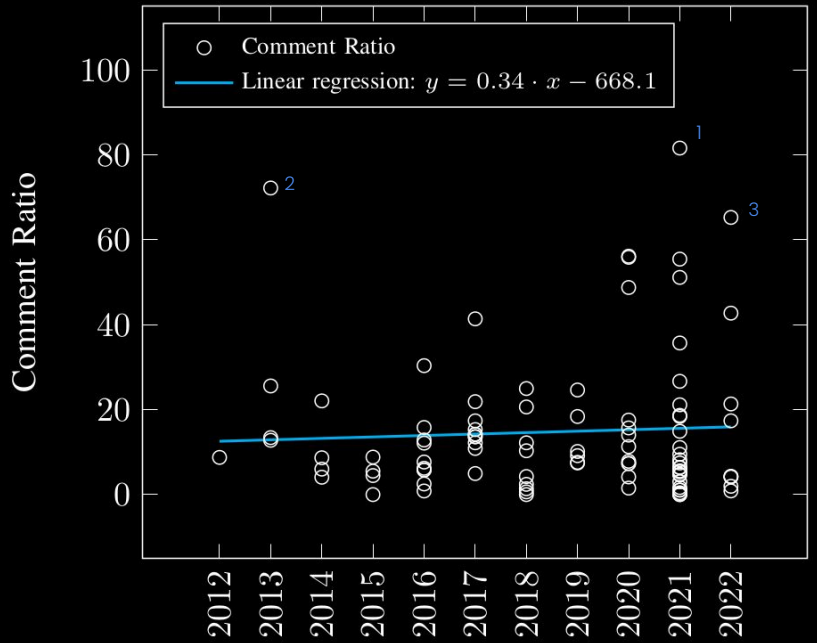
Maintainability By Year



1	Trojan	FakeFacebook	92.036
2	Trojan-SMS	MalRecipe	72.904
3	Rootkit	Adore	70.559

AVERAGE: 48.60

Density of Comments



1	Trojan-Spy	Flashlight	81.597
2	RAT	AndroidSurveillance	72.194
3	RAT	Gypte	65.265

AVERAGE: 14.81%

Android vs non-specific malware

Code Size, Development Cots, Code Quality

Limitations

- Non-representative dataset
  - Few samples
  - Collection bias
- Estimates, not reality
- Ever changing malware landscape
- UCC-J tool
- Too early

Conclusions

- Increase in code size
- Increase in development costs
- Decrease in code quality
- Larger sizes and costs than non-specific malware, fewer quality
- Inconclusive results

Future Work

Code reuse, clones and plagiarism

Malware vs Regular software

Compilation & Execution

More insights

Repeat in due time

Thanks !