

Malware Hunting with Machine Learning

Malware Threat Intelligence using a framework for multiple ML models.

by

Saket Upadhyay

About Saket Upadhyay



- Independent Security Researcher

About Saket Upadhyay



- Independent Security Researcher
- Interested in Malware Analysis, R.E., M.L.

About Saket Upadhyay



- ☛ Independent Security Researcher
- ☛ Interested in Malware Analysis, R.E., M.L. and everything in between.

About Saket Upadhyay



- Independent Security Researcher
- Interested in Malware Analysis, R.E., M.L. and everything in between.
- Worked in some companies as PenTester and Vulnerability Analyst

About Saket Upadhyay



- ☛ Independent Security Researcher
- ☛ Interested in Malware Analysis, R.E., M.L. and everything in between.
- ☛ Worked in some companies as PenTester and Vulnerability Analyst
- ☛ 5 years in the field of cybersecurity. (learning and research)

About Saket Upadhyay



- ☛ Independent Security Researcher
- ☛ Interested in Malware Analysis, R.E., M.L. and everything in between.
- ☛ Worked in some companies as PenTester and Vulnerability Analyst
- ☛ 5 years in the field of cybersecurity. (learning and research)
- ☛ Currently pursuing Bachelor's: Cybersec. and Digital Forensics @ VIT Bhopal

Contents

Contents

1. What is a malware?

How we define malware and what we understand from it.

Contents

1. What is a malware?

How we define malware and what we understand from it.

2. Basic Detection Methods.

Static, Dynamic and Hybrid Methods.

Contents

1. What is a malware?

How we define malware and what we understand from it.

2. Basic Detection Methods.

Static, Dynamic and Hybrid Methods.

3. Basic idea of Machine Learning in detection

Why we need ML? What we get from ML in detection?

Contents

1. What is a malware?

How we define malware and what we understand from it.

2. Basic Detection Methods.

Static, Dynamic and Hybrid Methods.

3. Basic idea of Machine Learning in detection

Why we need ML? What we get from ML in detection?

4. Feature Extraction of Malware Samples

What we look for in malware to train ML models?

Contents

1. What is a malware?

How we define malware and what we understand from it.

2. Basic Detection Methods.

Static, Dynamic and Hybrid Methods.

3. Basic idea of Machine Learning in detection

Why we need ML? What we get from ML in detection?

4. Feature Extraction of Malware Samples

What we look for in malware to train ML models?

5. Training and Testing Models

How we use extracted features for training?

Contents

1. What is a malware?

How we define malware and what we understand from it.

2. Basic Detection Methods.

Static, Dynamic and Hybrid Methods.

3. Basic idea of Machine Learning in detection

Why we need ML? What we get from ML in detection?

4. Feature Extraction of Malware Samples

What we look for in malware to train ML models?

5. Training and Testing Models

How we use extracted features for training?

6. Trying to Evade Single Feature Detection.

Let's try to bypass single feature single model detection.

Contents

1. What is a malware?

How we define malware and what we understand from it.

2. Basic Detection Methods.

Static, Dynamic and Hybrid Methods.

3. Basic idea of Machine Learning in detection

Why we need ML? What we get from ML in detection?

4. Feature Extraction of Malware Samples

What we look for in malware to train ML models?

5. Training and Testing Models

How we use extracted features for training?

6. Trying to Evade Single Feature Detection.

Let's try to bypass single feature single model detection.

7. Threat Intelligence & P.A.C.E. Framework

Combining multiple ML models to gather rich insight into malware trends.

Contents

1. What is a malware?

How we define malware and what we understand from it.

2. Basic Detection Methods.

Static, Dynamic and Hybrid Methods.

3. Basic idea of Machine Learning in detection

Why we need ML? What we get from ML in detection?

4. Feature Extraction of Malware Samples

What we look for in malware to train ML models?

5. Training and Testing Models

How we use extracted features for training?

6. Trying to Evade Single Feature Detection.

Let's try to bypass single feature single model detection.

7. Threat Intelligence & P.A.C.E. Framework

Combining multiple ML models to gather rich insight into malware trends.

8. Final Thoughts and Q& A

What is malware?

Textbook definition

Malware is any **software intentionally designed to cause damage** to a computer, server, client, or computer network (by contrast, software that causes *unintentional harm due to some deficiency is usually a software bug*).

Textbook definition

Malware is any **software intentionally designed to cause damage** to a computer, server, client, or computer network (by contrast, software that causes *unintentional harm due to some deficiency is usually a software bug*).

SOURCE : WIKIPEDIA

Textbook definition

A wide variety of types of malware exist, including

Textbook definition

A wide variety of types of malware exist, including

- Viruses ["Creeper system", 1971]

Textbook definition

A wide variety of types of malware exist, including

- ✖ Viruses ["Creeper system", 1971]
- ✖ Worms ["The Morris worm", 1988]

Textbook definition

A wide variety of types of malware exist, including

- ✖ Viruses ["Creeper system", 1971]
- ✖ Worms ["The Morris worm", 1988]
- ✖ Trojan horses ["ANIMAL", 1975]

Textbook definition

A wide variety of types of malware exist, including

- ▣ Viruses ["Creeper system", 1971]
- ▣ Worms ["The Morris worm", 1988]
- ▣ Trojan horses ["ANIMAL", 1975]
- ▣ Ransomware ["PC Cyborg", 1991]

Textbook definition

A wide variety of types of malware exist, including

- Viruses ["Creeper system", 1971]
- Worms ["The Morris worm", 1988]
- Trojan horses ["ANIMAL", 1975]
- Ransomware ["PC Cyborg", 1991]
- Spyware [around 2000]

Textbook definition

A wide variety of types of malware exist, including

- ☛ Viruses ["Creeper system", 1971]
- ☛ Worms ["The Morris worm", 1988]
- ☛ Trojan horses ["ANIMAL", 1975]
- ☛ Ransomware ["PC Cyborg", 1991]
- ☛ Spyware [around 2000]
- ☛ Adware [roughly in 1995]

Textbook definition

A wide variety of types of malware exist, including

- ☛ Viruses ["Creeper system", 1971]
- ☛ Worms ["The Morris worm", 1988]
- ☛ Trojan horses ["ANIMAL", 1975]
- ☛ Ransomware ["PC Cyborg", 1991]
- ☛ Spyware [around 2000]
- ☛ Adware [roughly in 1995]
- ☛ Scareware. [increase in activity since 2008]

Key behaviour observations

Key behaviour observations

- ✖ Stealing [copy, move and]

Key behaviour observations

- ✖ Stealing [copy, move and]
- ✖ Encrypting data [file modification calls]

Key behaviour observations

- ✖ Stealing [copy, move and]
- ✖ Encrypting data [file modification calls]
- ✖ Deleting sensitive data [file manipulation/mod.]

Key behaviour observations

- ✖ Stealing [copy, move and]
- ✖ Encrypting data [file modification calls]
- ✖ Deleting sensitive data [file manipulation/mod.]
- ✖ Altering or hijacking core computing funct. [system calls]

Key behaviour observations

- Stealing [copy, move and]
- Encrypting data [file modification calls]
- Deleting sensitive data [file manipulation/mod.]
- Altering or hijacking core computing funct. [system calls]
- Monitoring users [active processes, process hooks etc.]

Key behaviour observations

- ☛ Stealing [copy, move and]
- ☛ Encrypting data [file modification calls]
- ☛ Deleting sensitive data [file manipulation/mod.]
- ☛ Altering or hijacking core computing funct. [system calls]
- ☛ Monitoring users [active processes, process hooks etc.]
- ☛ Process hijack[memory injection]
- ☛ Et Cetera

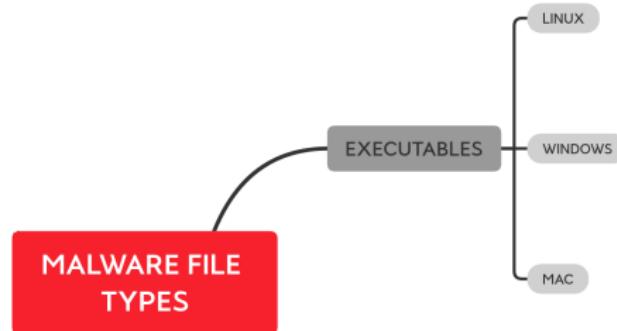
Malware Infection Vectors

MALWARE FILE
TYPES

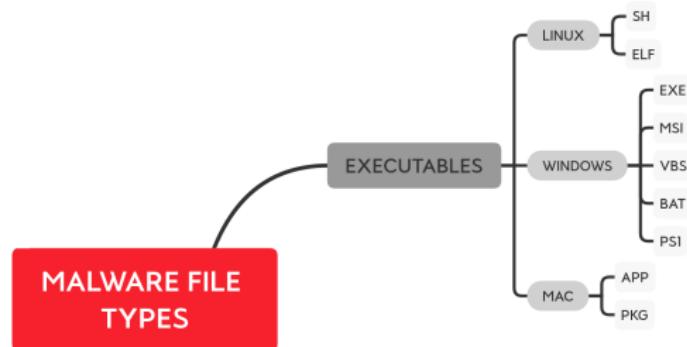
Malware Infection Vectors



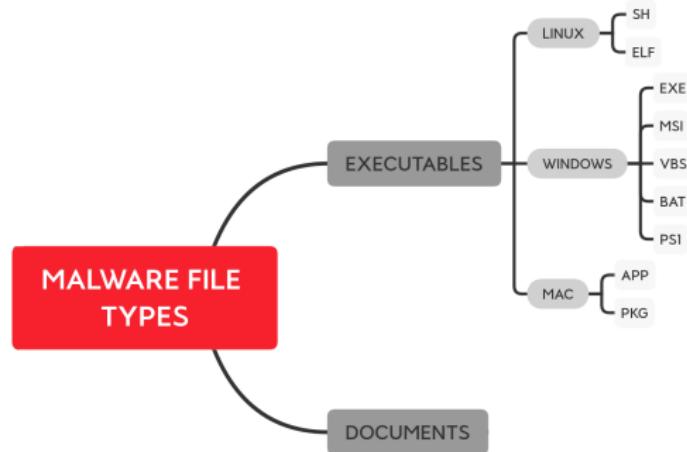
Malware Infection Vectors



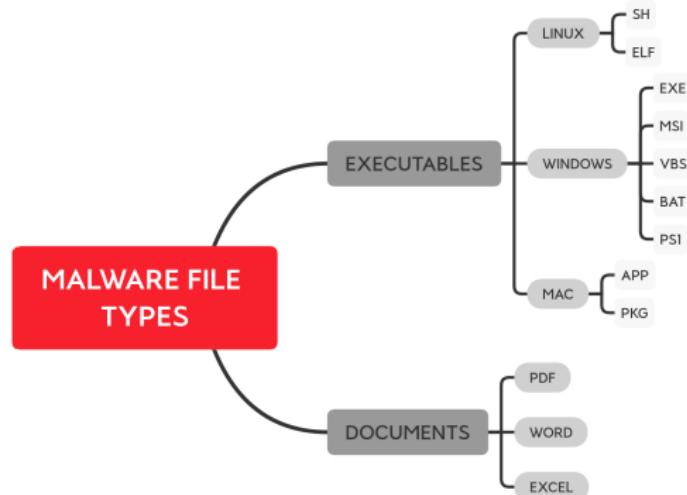
Malware Infection Vectors



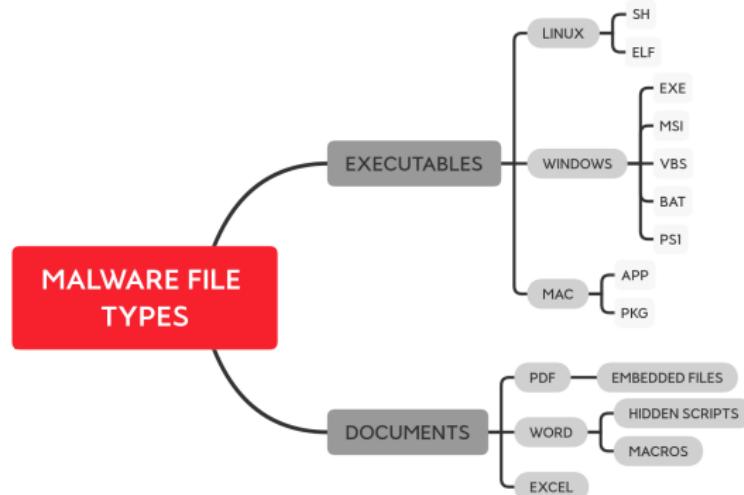
Malware Infection Vectors



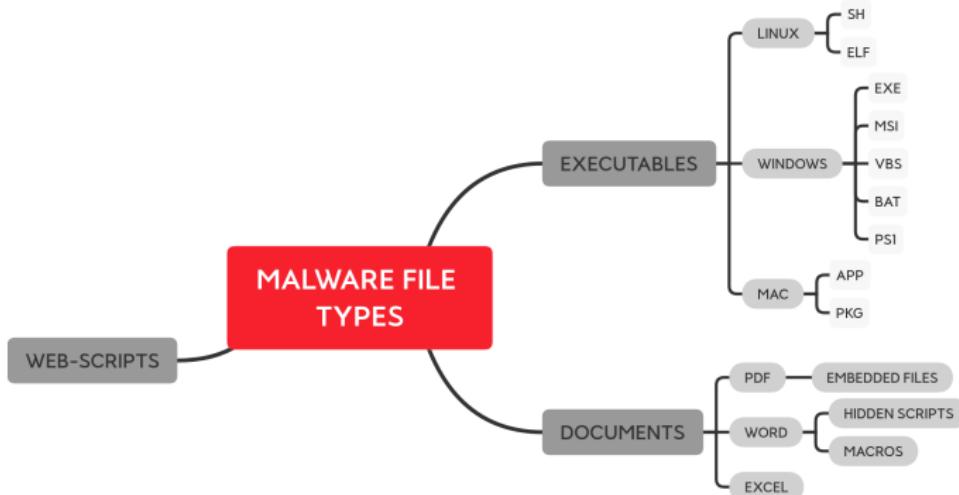
Malware Infection Vectors



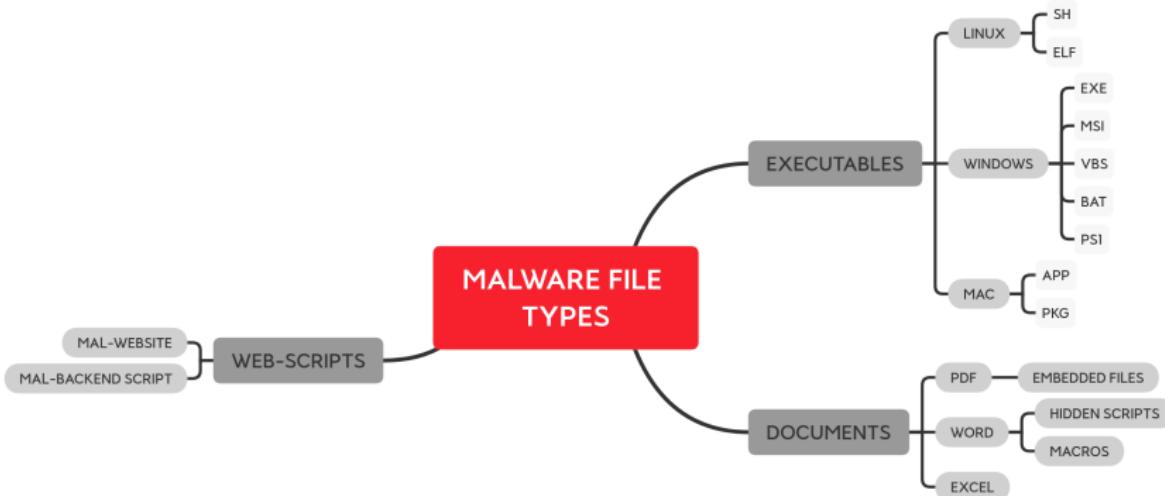
Malware Infection Vectors



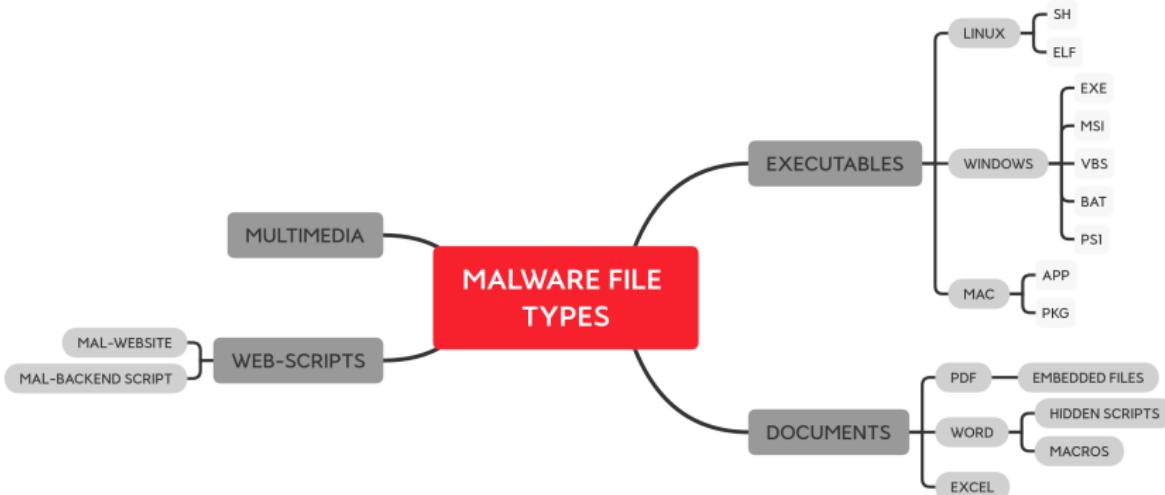
Malware Infection Vectors



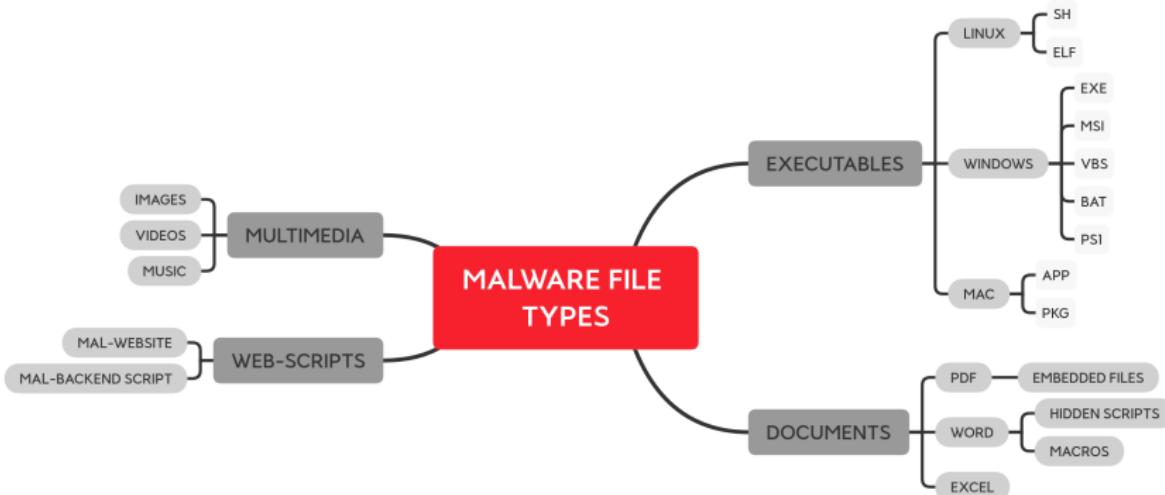
Malware Infection Vectors



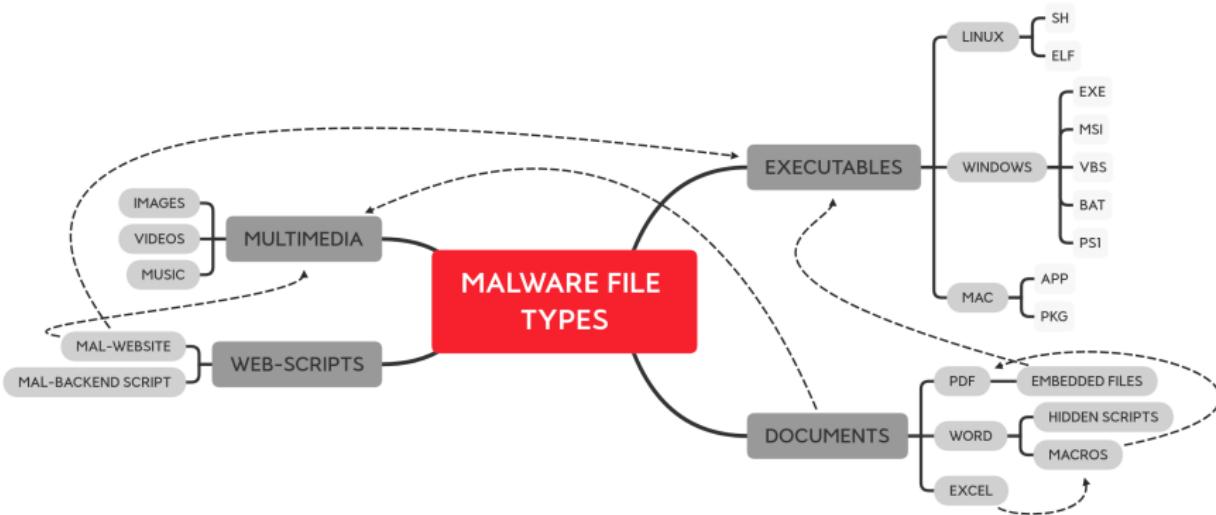
Malware Infection Vectors



Malware Infection Vectors



Malware Infection Vectors



Basic Detection Methods

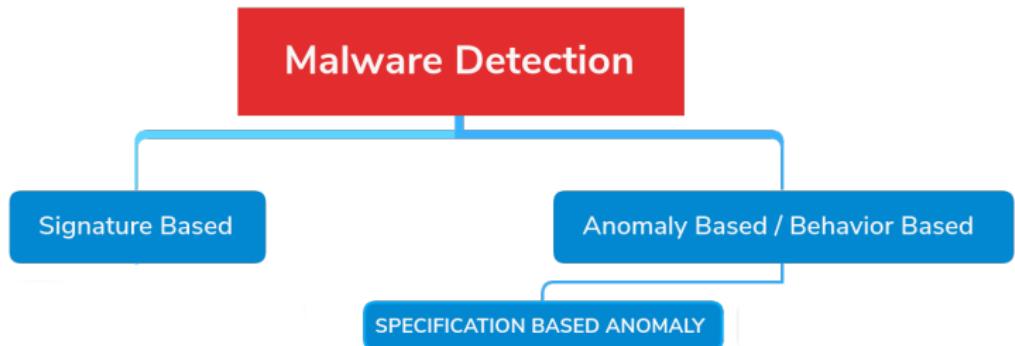
Detection Strategies



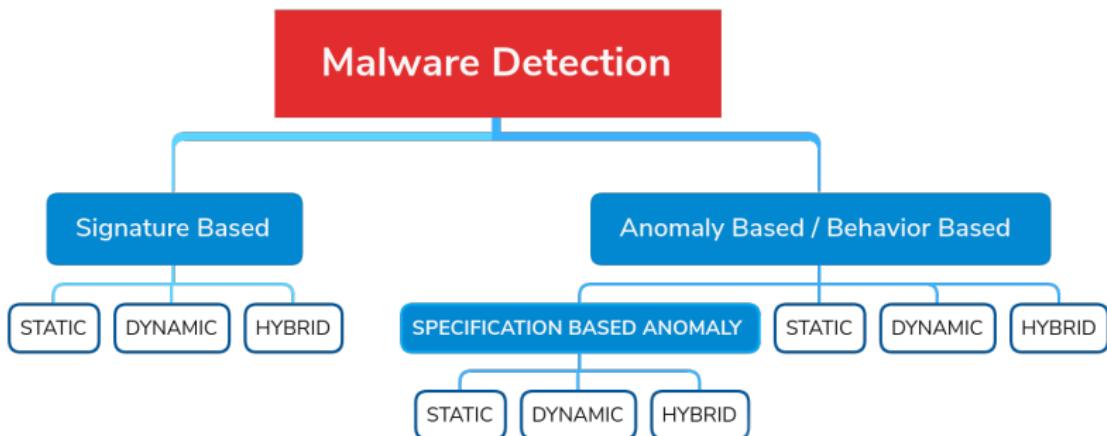
Detection Strategies



Detection Strategies



Detection Strategies



STATIC & DYNAMIC Analysis

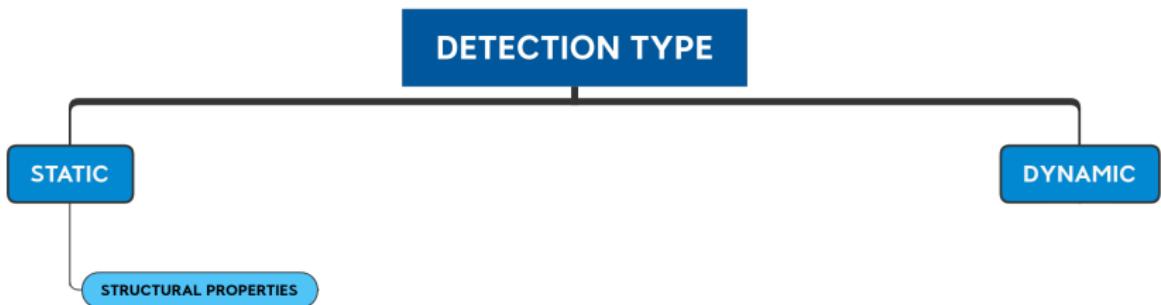
DETECTION TYPE

STATIC

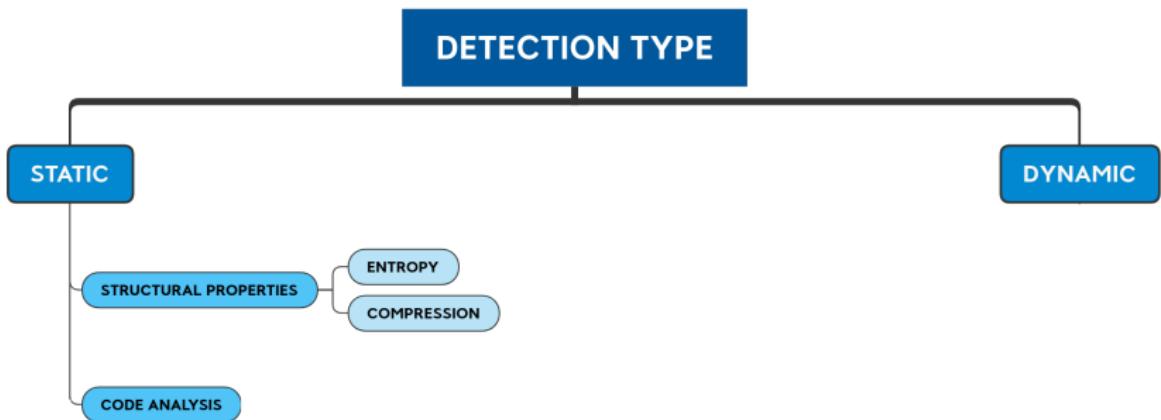
STATIC & DYNAMIC Analysis



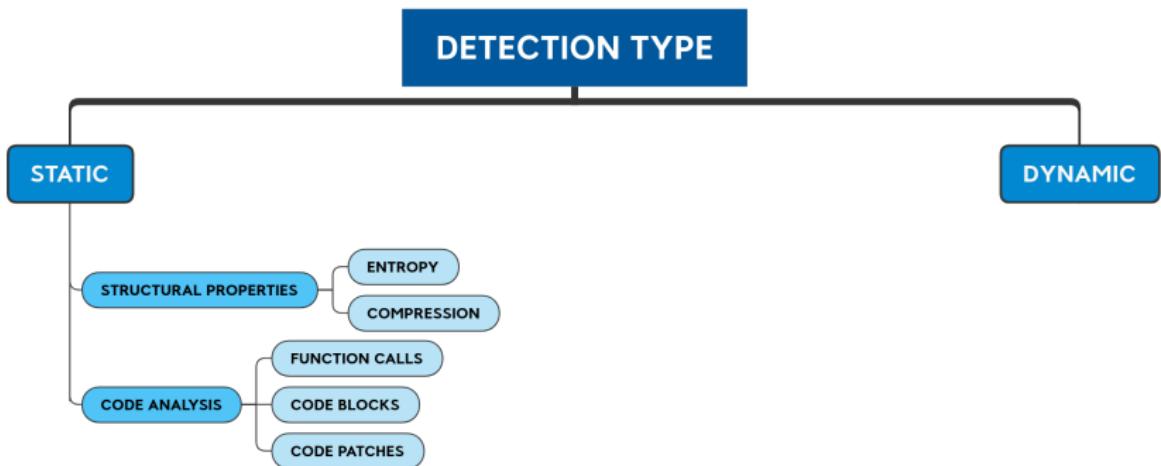
STATIC & DYNAMIC Analysis



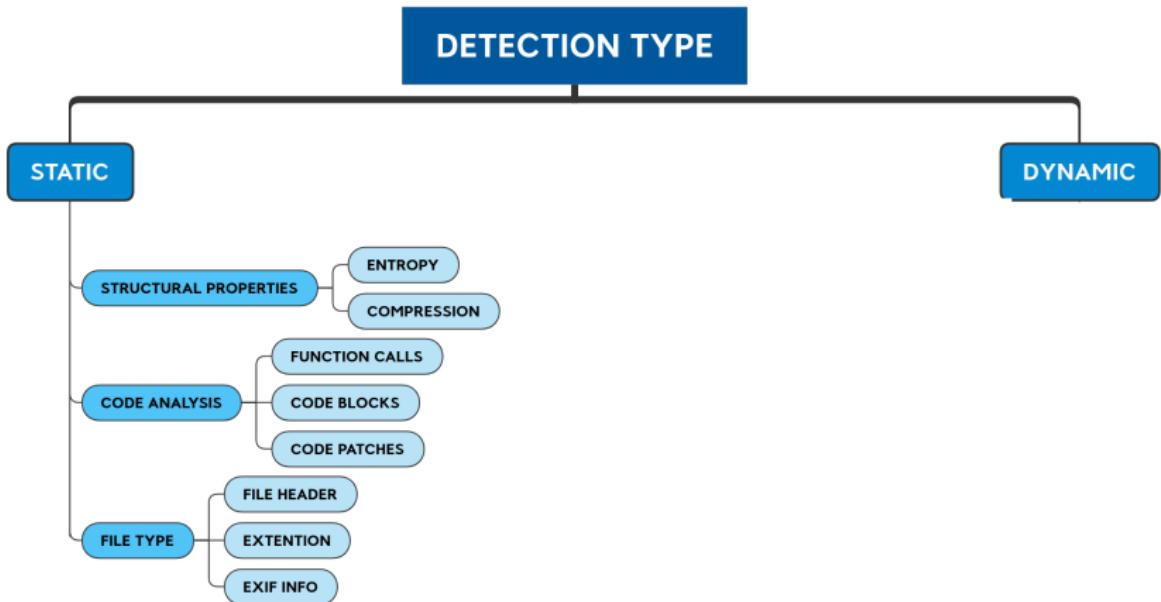
STATIC & DYNAMIC Analysis



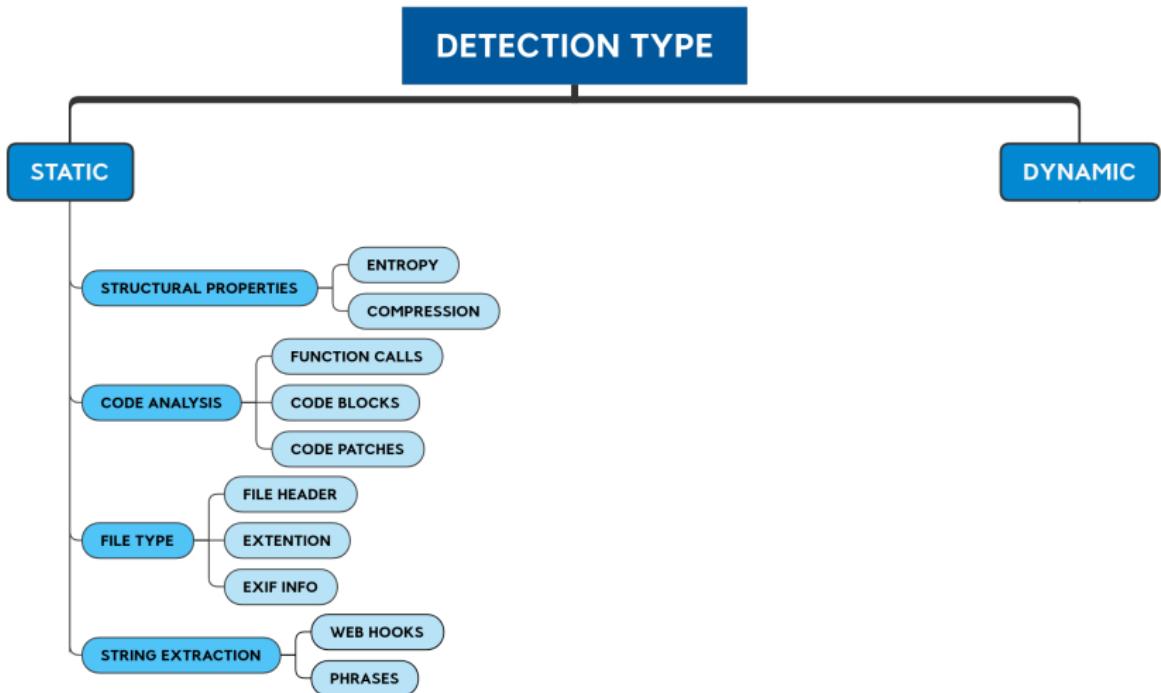
STATIC & DYNAMIC Analysis



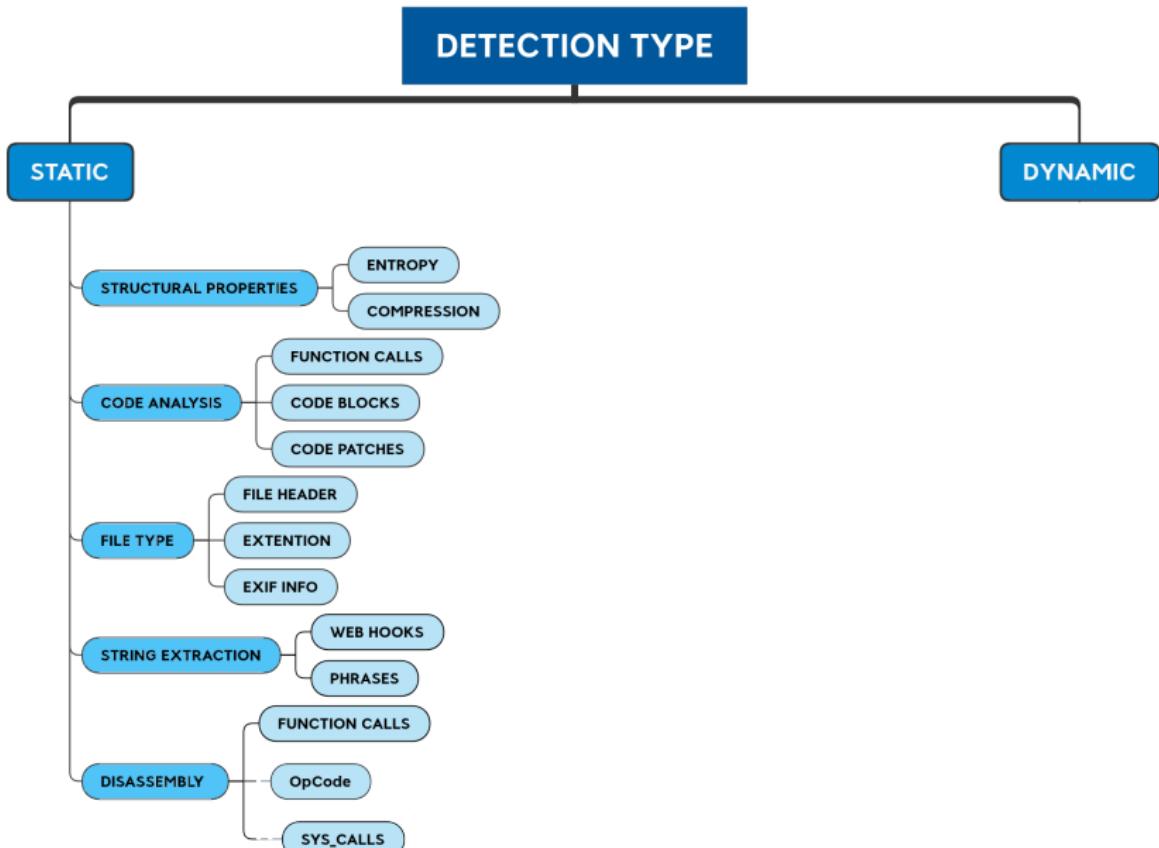
STATIC & DYNAMIC Analysis



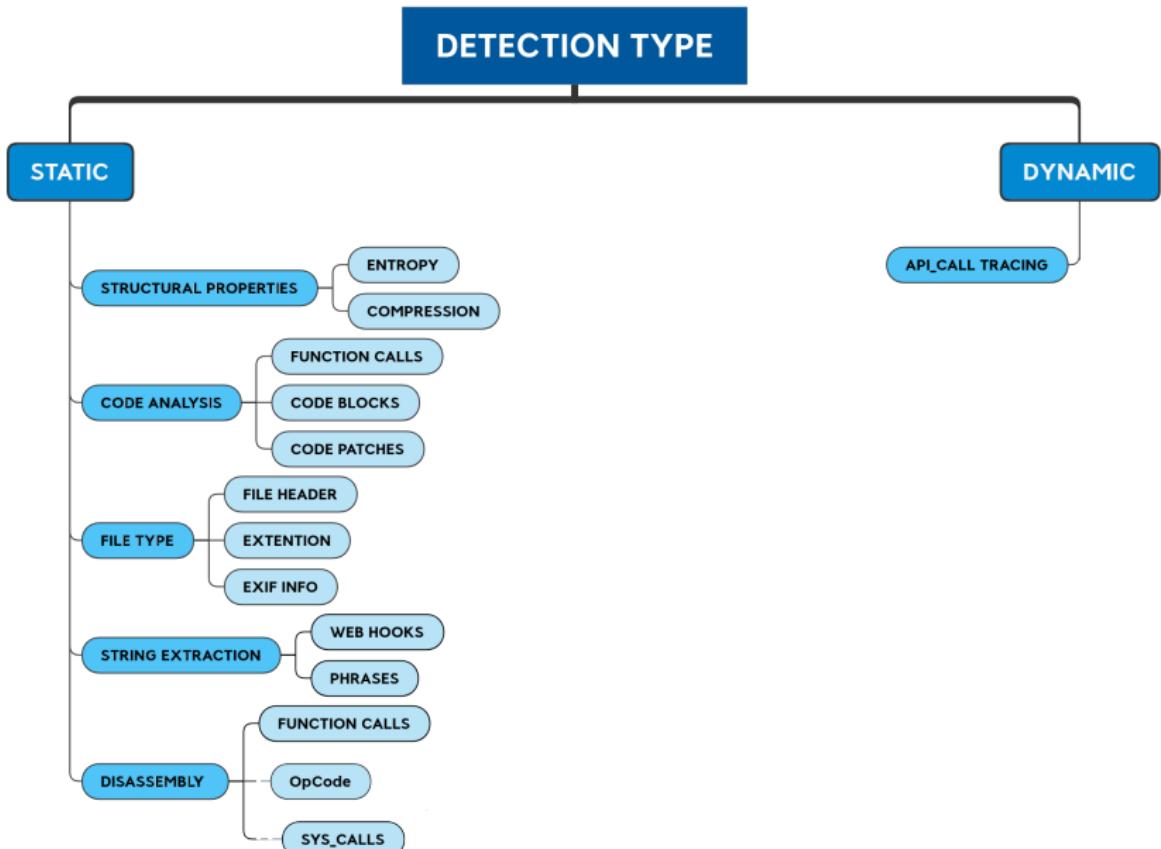
STATIC & DYNAMIC Analysis



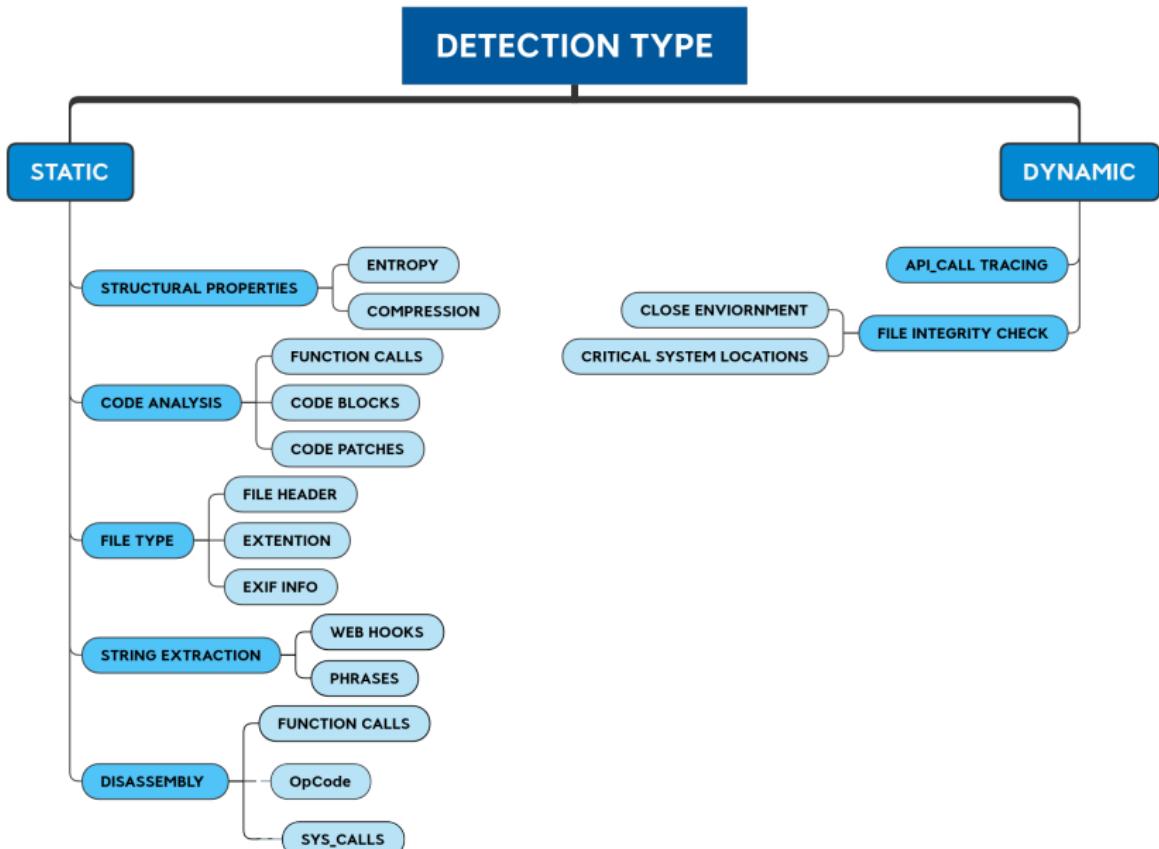
STATIC & DYNAMIC Analysis



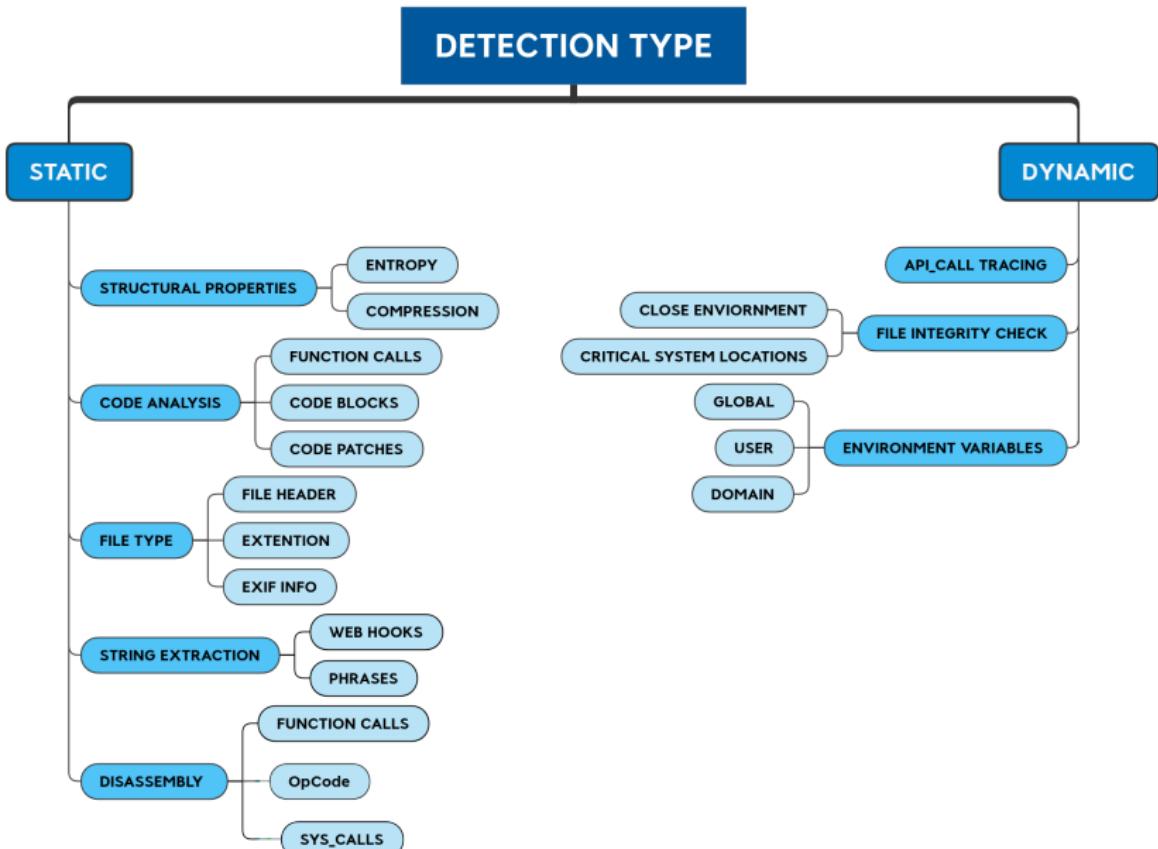
STATIC & DYNAMIC Analysis



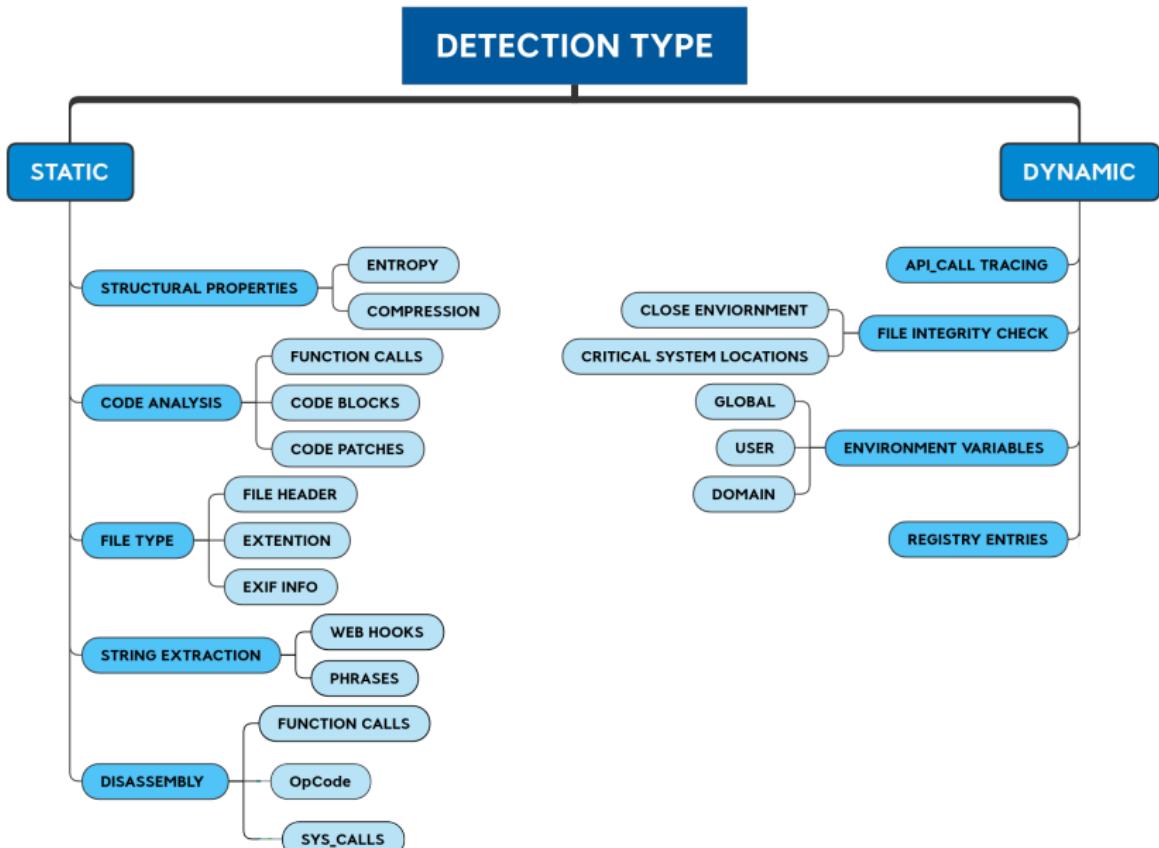
STATIC & DYNAMIC Analysis



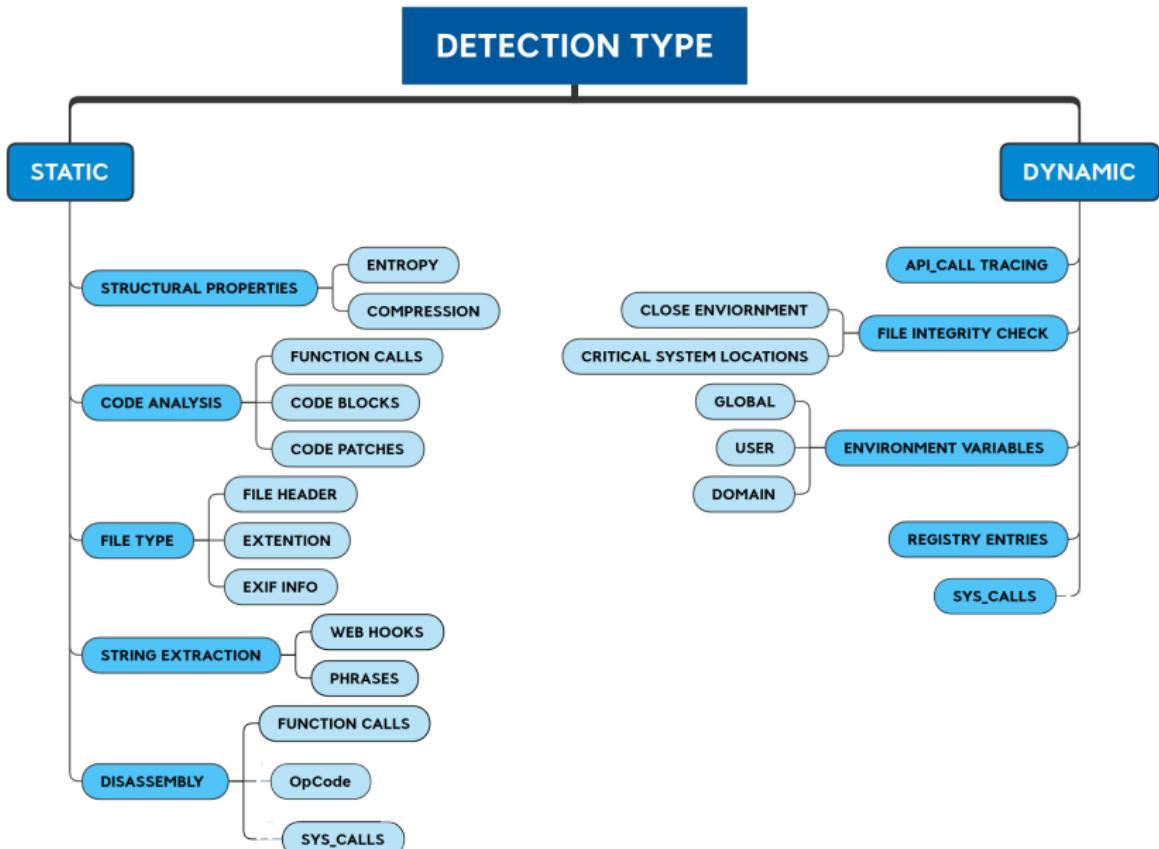
STATIC & DYNAMIC Analysis



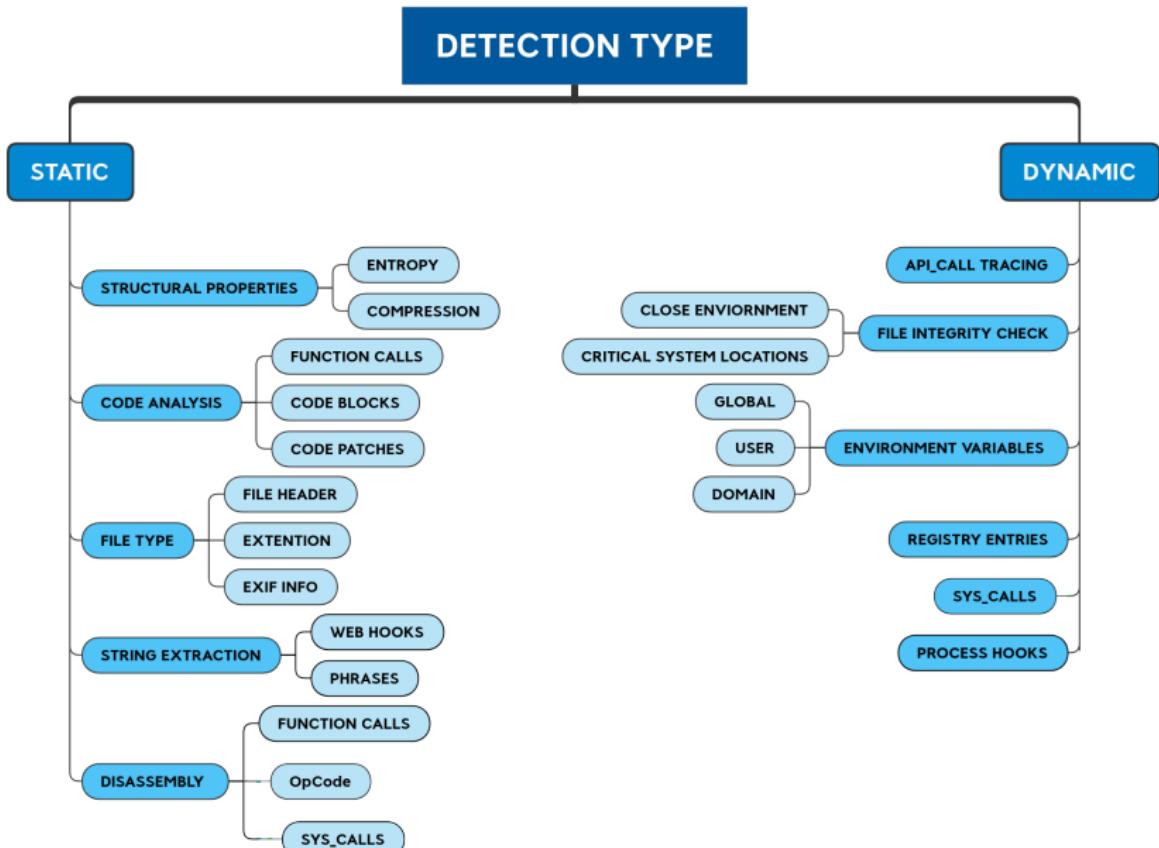
STATIC & DYNAMIC Analysis



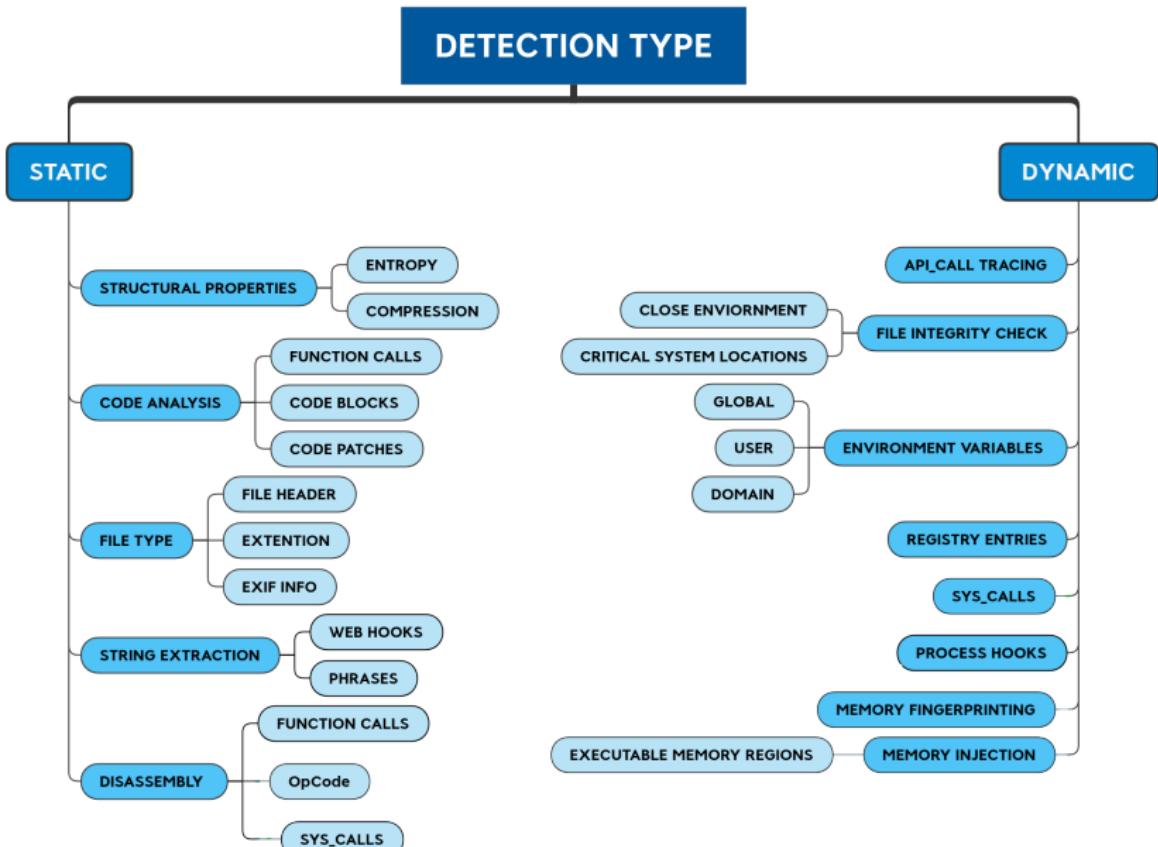
STATIC & DYNAMIC Analysis



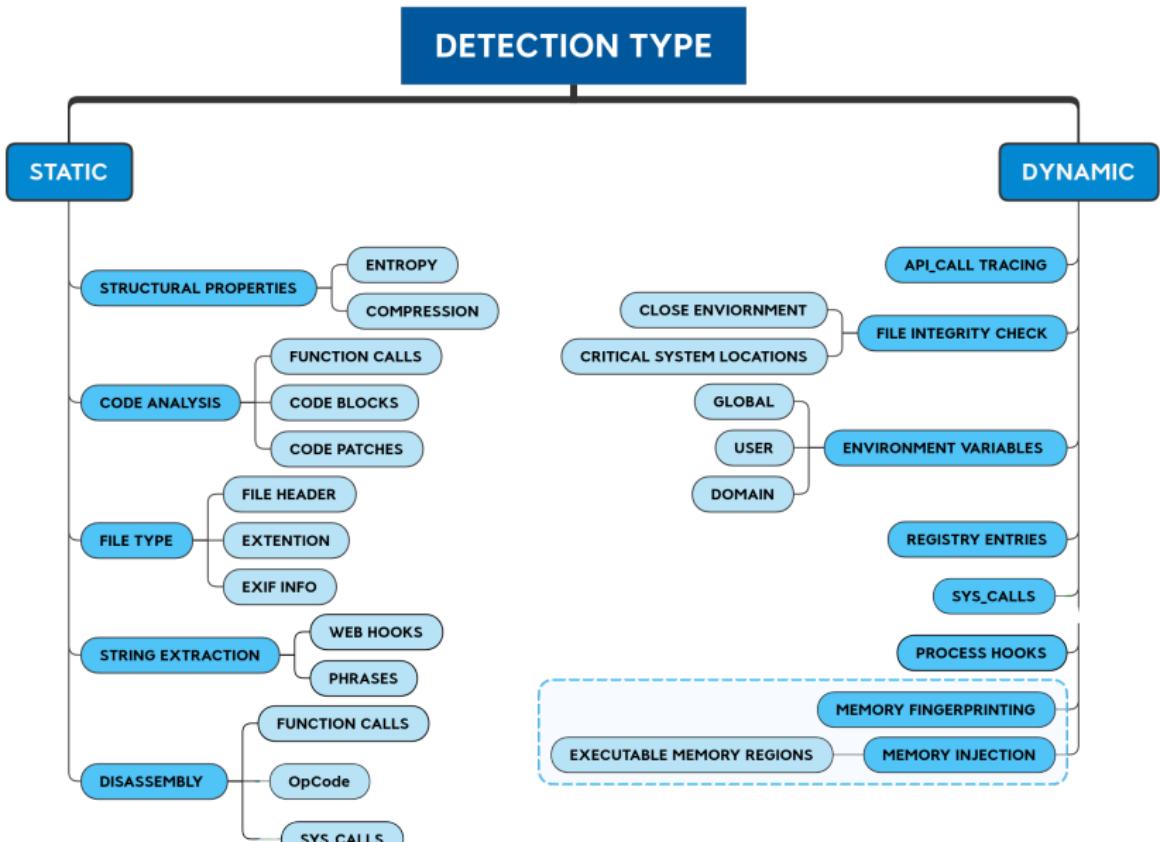
STATIC & DYNAMIC Analysis



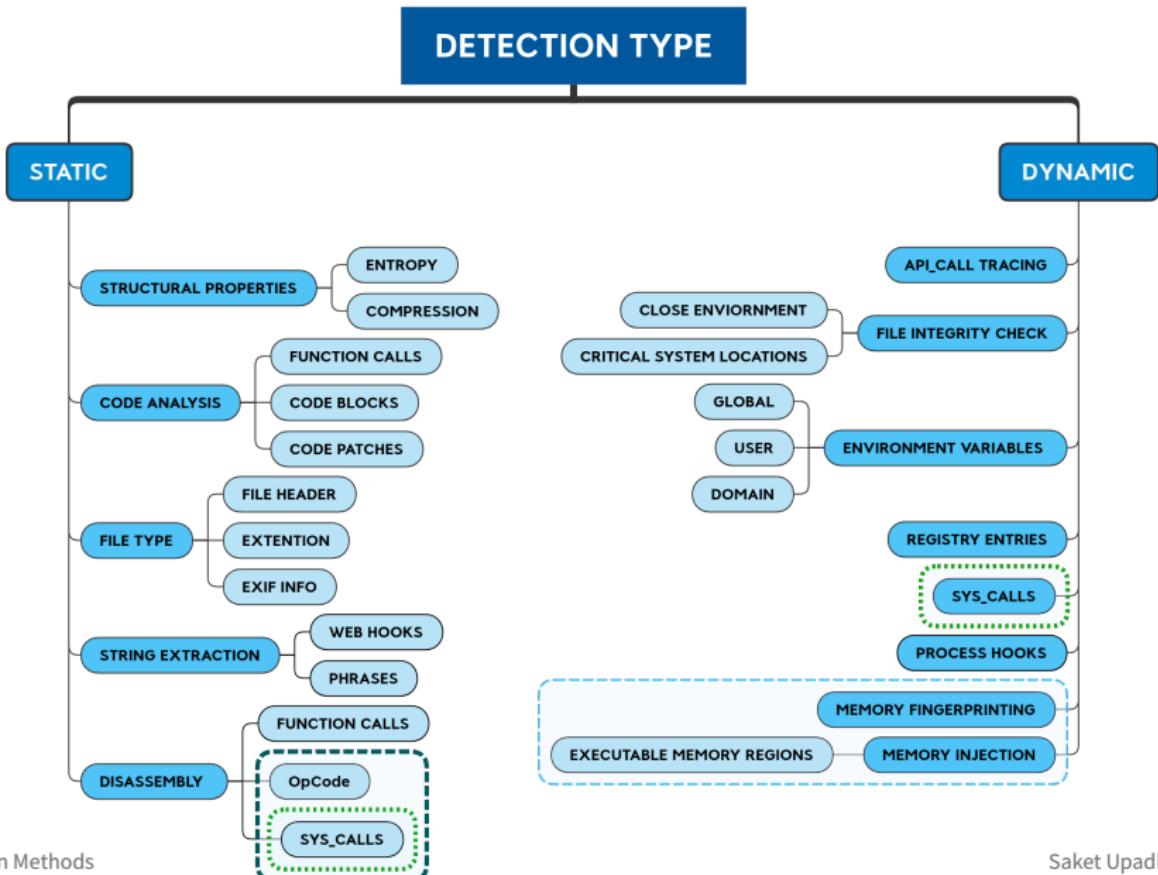
STATIC & DYNAMIC Analysis



STATIC & DYNAMIC Analysis



STATIC & DYNAMIC Analysis



Advantages & Disadvantages



SIGNATURE BASED

Advantages & Disadvantages



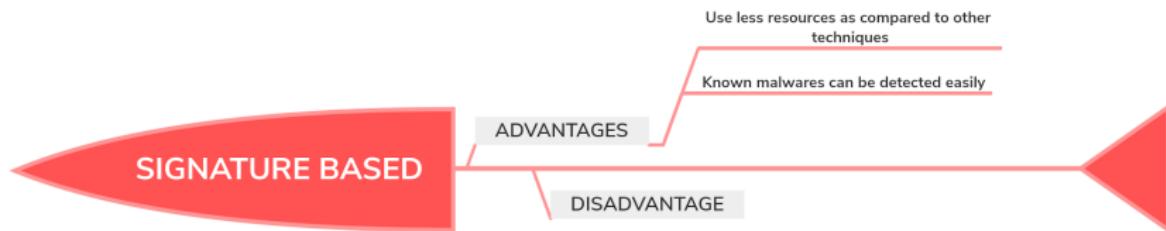
Advantages & Disadvantages



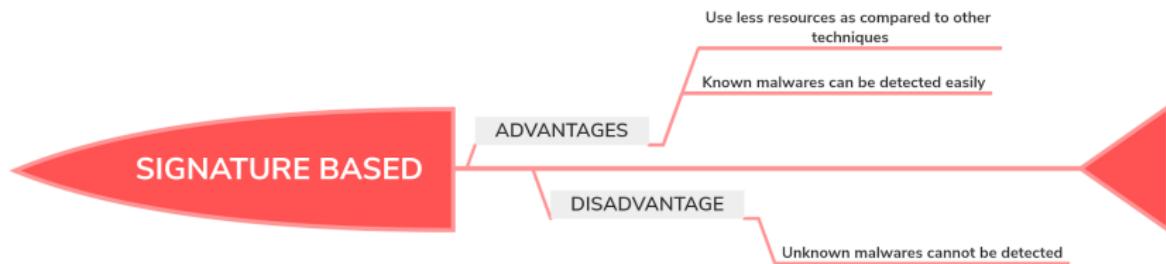
Advantages & Disadvantages



Advantages & Disadvantages



Advantages & Disadvantages



Advantages & Disadvantages



HEURISTIC BASED

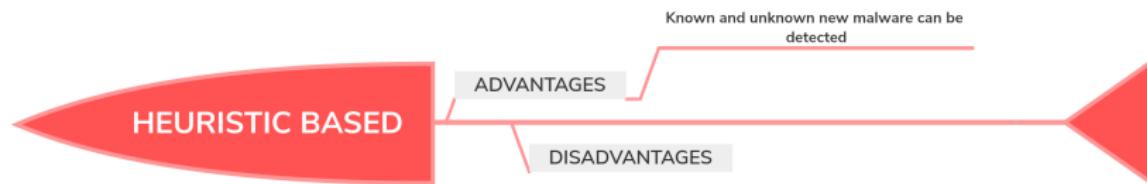
Advantages & Disadvantages



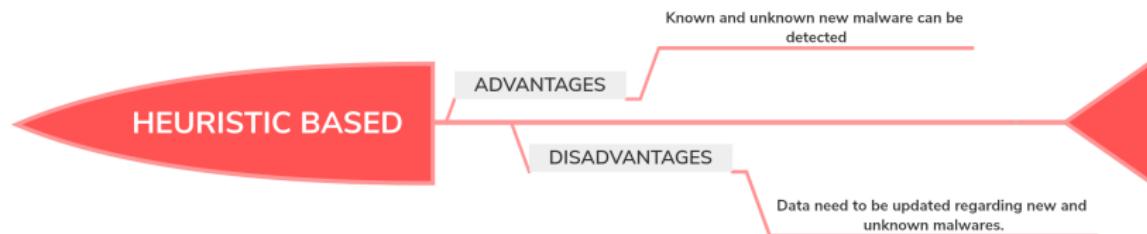
Advantages & Disadvantages



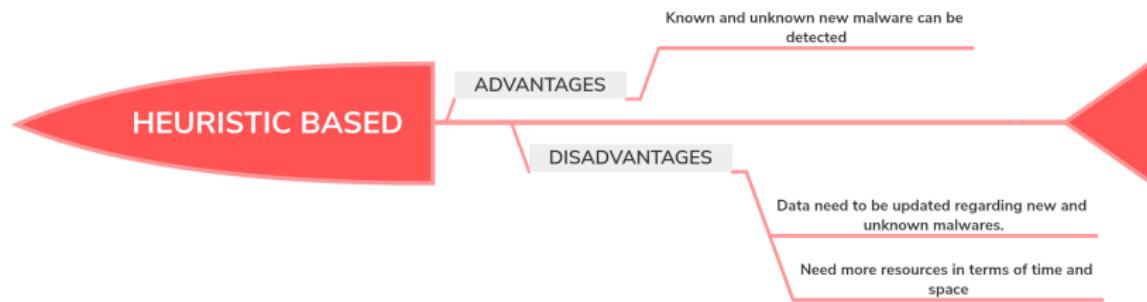
Advantages & Disadvantages



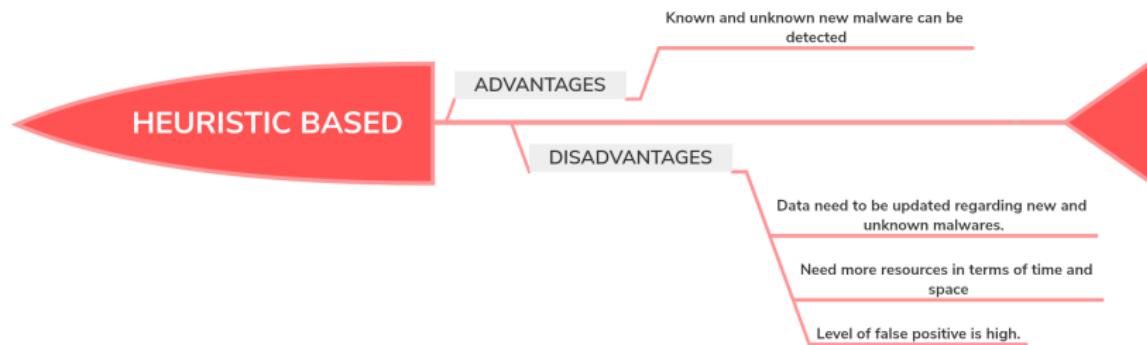
Advantages & Disadvantages



Advantages & Disadvantages

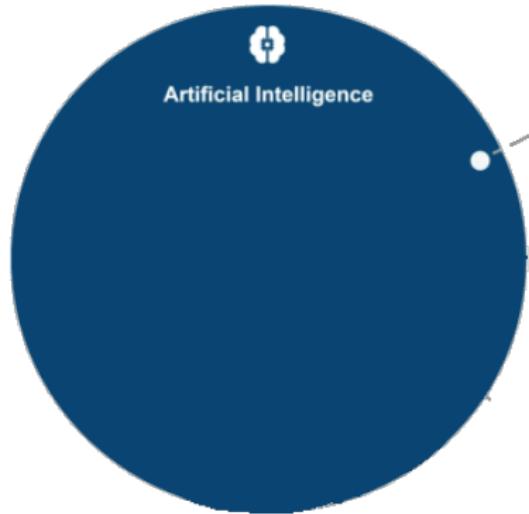


Advantages & Disadvantages



Basic Idea of Machine Learning in detection

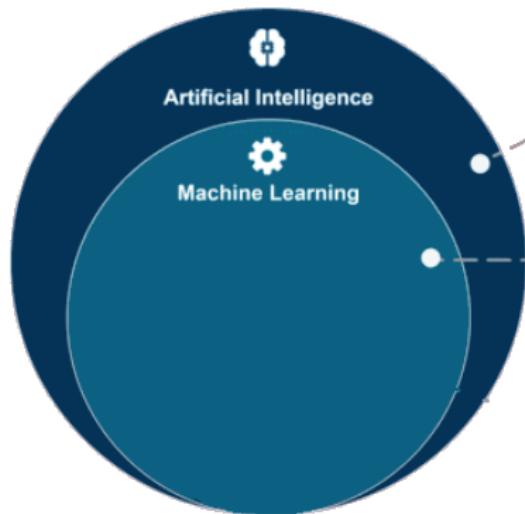
Artificial Intelligence



ARTIFICIAL INTELLIGENCE

A technique which enables machines
to mimic human behaviour

Artificial Intelligence



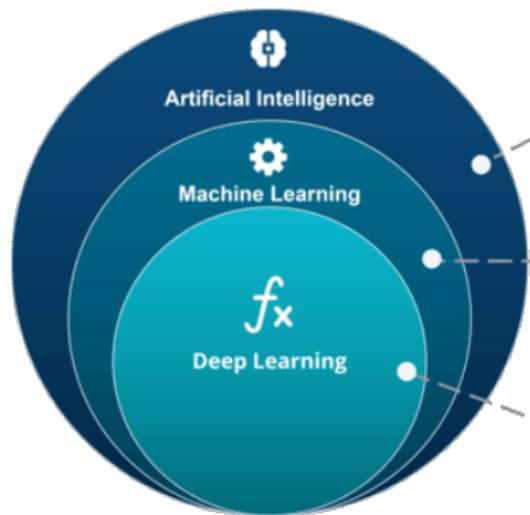
ARTIFICIAL INTELLIGENCE

A technique which enables machines to mimic human behaviour

MACHINE LEARNING

A subset of AI technique which uses statistical methods to enable machines to improve with experience

Artificial Intelligence



ARTIFICIAL INTELLIGENCE

A technique which enables machines to mimic human behaviour

MACHINE LEARNING

A subset of AI technique which use statistical methods to enable machines to improve with experience

DEEP LEARNING

A subset of ML which make the computation of multi-layer neural network feasible

What we will talk about ?

Artificial Intelligence

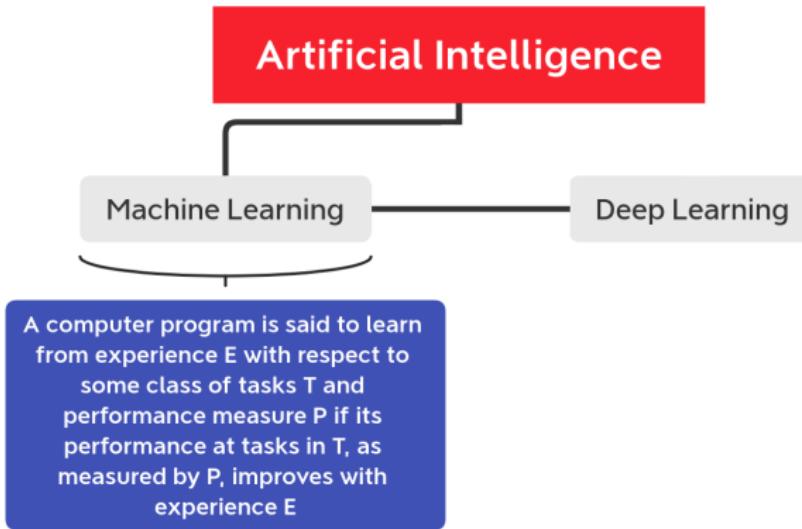
What we will talk about ?



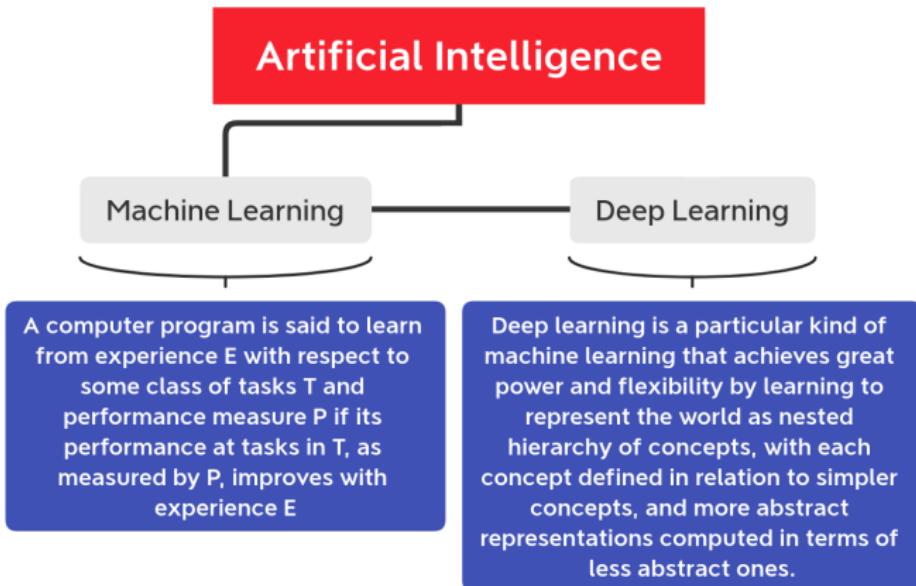
What we will talk about ?



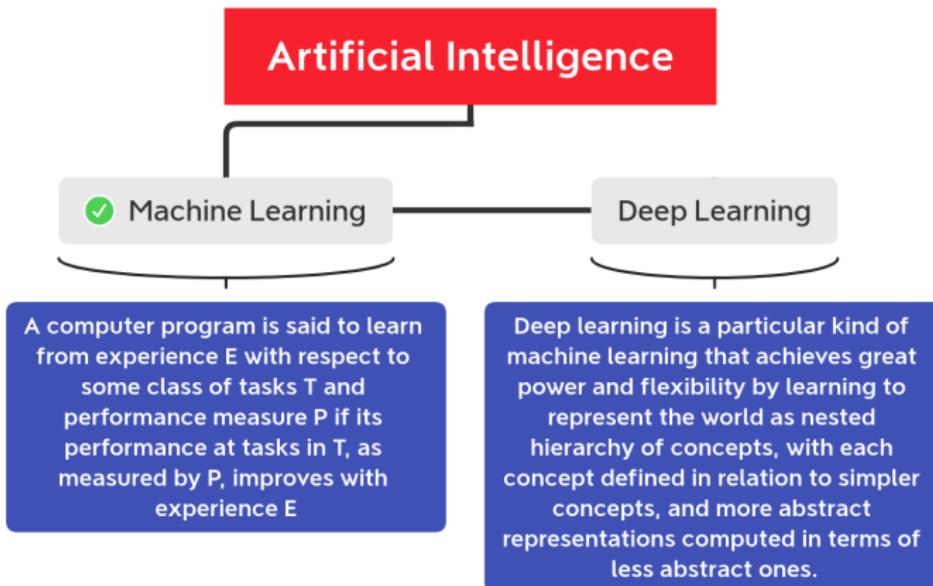
What we will talk about ?



What we will talk about ?



What we will talk about ?



Use of ML and DL

Scenario 1:

We have to build a surveillance system which will scan the faces of people and raise an alarm if not recognised.

Use of ML and DL

Scenario 1:

We have to build a surveillance system which will scan the faces of people and raise an alarm if not recognised.

Scenario 2:

Given a bank's transaction details and credit transfers history we have to detect fraud transactions.

Use of ML and DL

Scenario 1:

We have to build a surveillance system which will scan the faces of people and raise an alarm if not recognised. **[DEEP LEARNING]**

Scenario 2:

Given a bank's transaction details and credit transfers history we have to detect fraud transactions. **[MACHINE LEARNING]**

Classification

In machine learning and statistics, classification is the problem of **identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set** of data containing observations (or instances) whose category membership is **known**.

Classification

In machine learning and statistics, classification is the problem of **identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set** of data containing observations (or instances) whose category membership is **known**.

source : wikipedia

Classification Algorithms in ML

Classification Algorithms in ML

- ✚ Linear Classifiers: Logistic Regression, Naive Bayes Classifier

Classification Algorithms in ML

- ✚ Linear Classifiers: Logistic Regression, Naive Bayes Classifier
- ✚ Nearest Neighbor

Classification Algorithms in ML

- ✚ Linear Classifiers: Logistic Regression, Naive Bayes Classifier
- ✚ Nearest Neighbor
- ✚ Support Vector Machines

Classification Algorithms in ML

- ✚ Linear Classifiers: Logistic Regression, Naive Bayes Classifier
- ✚ Nearest Neighbor
- ✚ Support Vector Machines
- ✚ Decision Trees

Classification Algorithms in ML

- ✚ Linear Classifiers: Logistic Regression, Naive Bayes Classifier
- ✚ Nearest Neighbor
- ✚ Support Vector Machines
- ✚ Decision Trees
- ✚ Boosted Trees

Classification Algorithms in ML

- ✚ Linear Classifiers: Logistic Regression, Naive Bayes Classifier
- ✚ Nearest Neighbor
- ✚ Support Vector Machines
- ✚ Decision Trees
- ✚ Boosted Trees
- ✚ Random Forest

Classification Algorithms in ML

- ✚ Linear Classifiers: Logistic Regression, Naive Bayes Classifier
- ✚ Nearest Neighbor
- ✚ Support Vector Machines
- ✚ Decision Trees
- ✚ Boosted Trees
- ✚ Random Forest
- ✚ Neural Networks

Visualisation of SVM

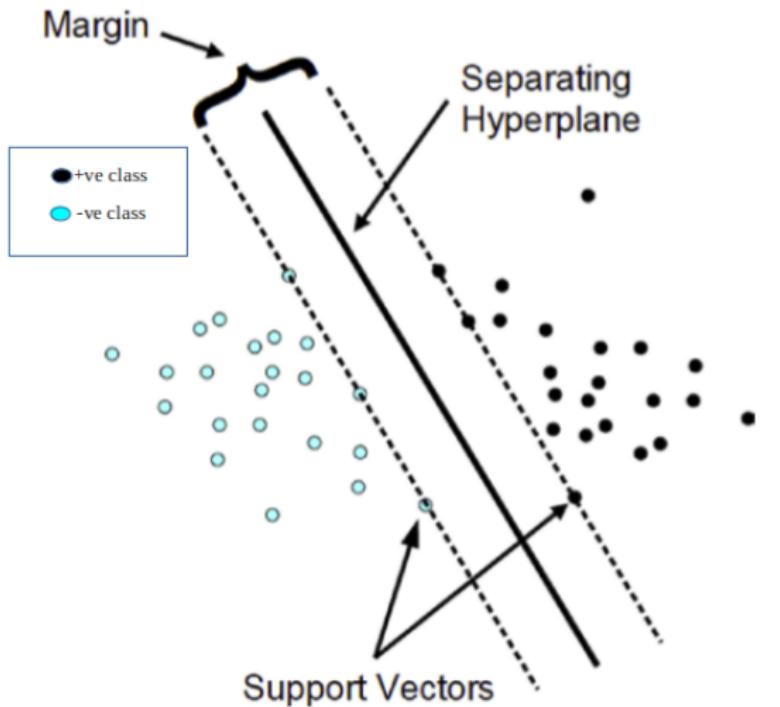


Fig 1:SVM Visualisation

Visualisation of KNN



Fig 2:KNN Visualisation

Steps to follow ...

1. Collect Samples of malware and benign applications

Steps to follow ...

1. Collect Samples of malware and benign applications
2. Feature engineering

Steps to follow ...

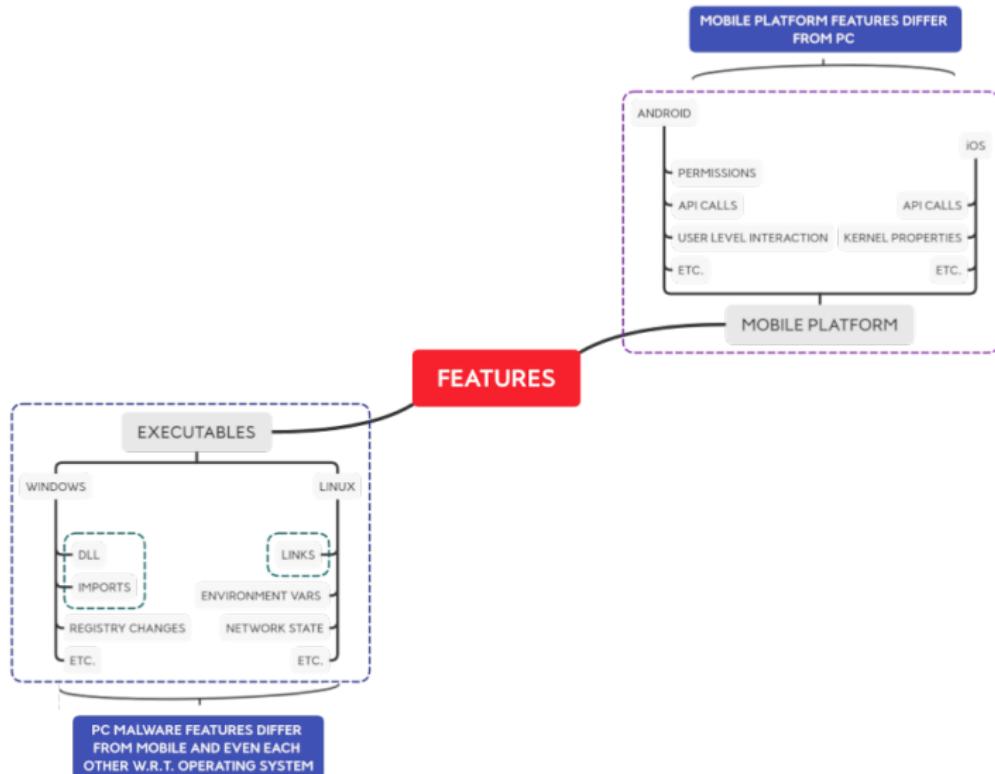
1. Collect Samples of malware and benign applications
2. Feature engineering
3. Training and testing

Steps to follow ...

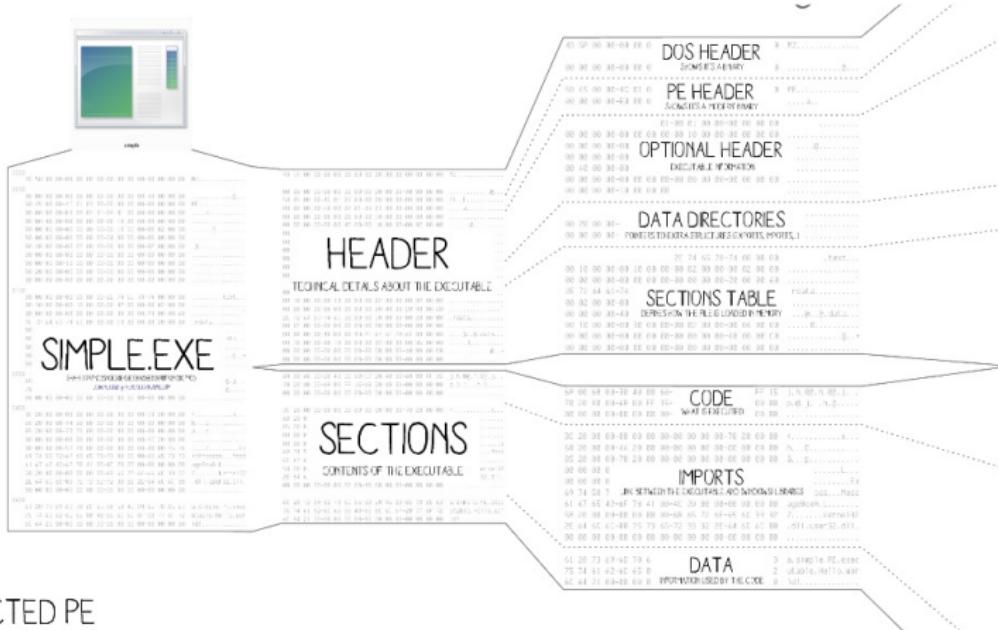
1. Collect Samples of malware and benign applications
2. Feature engineering
3. Training and testing
4. Deployment

Feature Extraction of Malware Samples

What we should look for in malware?



PE File Format



src : <https://www.slideshare.net/ange4771/44con2013-workshop-exploring-the-portable-executable-format>

Feature Extraction of Malware Samples

Saket Upadhyay

PE File Format

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00000000	4D	5A	90	00	03	00	00	00	04	00	00	00	FF	FF	00	00
00000010	B8	00	00	00	00	00	00	40	00	00	00	00	00	00	00	00
00000020	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000030	00	00	00	00	00	00	00	00	00	00	00	00	80	00	00	00
00000040	0E	1F	BA	OE	00	B4	09	CD	21	B8	01	4C	CD	21	54	68
00000050	69	73	20	70	72	6F	67	72	61	6D	20	63	61	6E	6E	6F
00000060	74	20	62	65	20	72	75	6E	20	69	6E	20	44	4F	53	20
00000070	6D	6F	64	65	2E	0D	0D	0A	24	00	00	00	00	00	00	00
00000080	50	45	00	00	4C	01	03	00	8D	FA	81	4D	00	00	00	00
00000090	00	00	00	00	E0	00	02	01	OB	01	08	00	00	0A	00	00
000000A0	00	08	00	00	00	00	00	00	9E	28	00	00	00	20	00	00
000000B0	00	40	00	00	00	00	40	00	00	20	00	00	00	02	00	00
000000C0	04	00	00	00	00	00	00	00	04	00	00	00	00	00	00	00
000000D0	00	80	00	00	00	02	00	00	01	82	00	00	03	00	40	85
000000E0	00	00	10	00	00	10	00	00	00	00	10	00	00	10	00	00
000000F0	00	00	00	00	10	00	00	00	00	00	00	00	00	00	00	00
00000100	4C	28	00	00	4F	00	00	00	00	40	00	00	A8	05	00	00
00000110	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000120	00	60	00	00	0C	00	00	00	A4	27	00	00	1C	00	00	00
00000130	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000140	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

src : <https://www.red-gate.com/simple-talk/blogs/anatomy-of-a-net-assembly-pe-headers/>

PE File Format

Offset (h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00000000	4D	5A	90	00	03	00	00	00	04	00	00	00	FF	FF	00	00
00000010	B8	00	00	00	00	00	00	40	00	00	00	00	00	00	00	00
00000020	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000030	00	00	00	00	00	00	00	00	00	00	00	00	80	00	00	00
00000040	0E	1F	BA	OE	00	B4	09	CD	21	B8	01	4C	CD	21	54	68
00000050	69	73	20	70	72	6F	67	72	61	6D	20	63	61	6E	6E	6F
00000060	74	20	62	65	20	72	75	6E	20	69	6E	20	44	4F	53	20
00000070	6D	6F	64	65	2E	0D	0D	0A	24	00	00	00	00	00	00	00
00000080	50	45	00	00	4C	01	03	00	8D	FA	81	4D	00	00	00	00
00000090	00	00	00	00	E0	00	02	01	0B	01	08	00	00	0A	00	00
000000A0	00	08	00	00	00	00	00	00	9E	28	00	00	00	20	00	00
000000B0	00	40	00	00	00	00	40	00	00	20	00	00	00	02	00	00
000000C0	04	00	00	00	00	00	00	00	04	00	00	00	00	00	00	00
000000D0	00	80	00	00	00	02	00	00	01	82	00	00	03	00	40	85
000000E0	00	00	10	00	00	10	00	00	00	00	10	00	00	10	00	00
000000F0	00	00	00	00	10	00	00	00	00	00	00	00	00	00	00	00
00000100	4C	28	00	00	4F	00	00	00	00	40	00	00	A8	05	00	00
00000110	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000120	00	60	00	00	0C	00	00	00	A4	27	00	00	1C	00	00	00
00000130	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
00000140	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00

Source: https://www.tutorialspoint.com/pe_file_format.htm
Note: This is a simplified representation of PE file headers. In reality, there are many more fields and they are not assembly-like.

How we can store the features?

Now we know how and what to extract,

How we can store the features?

Now we know how and what to extract,

BUT

we need a well defined structure to store the features so that we
can feed them in our ML training step

The Binary Array Approach [1/0]

STEPS

The Binary Array Approach [1/0]

STEPS

1. We create a global set of features

The Binary Array Approach [1/0]

STEPS

1. We create a global set of features
2. Take one sample at a time

The Binary Array Approach [1/0]

STEPS

1. We create a global set of features
2. Take one sample at a time
3. If it's features exist in the global set, toggle value to '1' otherwise '0'

The Binary Array Approach [1/0]

STEPS

1. We create a global set of features
2. Take one sample at a time
3. If it's features exist in the global set, toggle value to '1' otherwise '0'
4. Repeat 2-3 for all samples

Feature Extraction DEMO

saketupadhyay@SPIDER: ~/PycharmProjects/Apk_Rever

Training and Testing Models

Why divide the dataset into parts?

	Purpose	Yield	Used for Model training	Used for Parameter tuning
Train Data	To learn patterns from the data.	A model that makes near-expected predictions	Yes	Yes
Validation Data	To understand model behaviour and generalizability on unseen data.	Insights on how to tune your model.	No	Yes
Test Data	To understand how the model would perform in real world scenario.	A completely unbiased estimate of model performance.	No	No

Table 1: Test Train Split

src : medium.com/datadriveninvestor/data-science-essentials-why-train-validation-test-data-b7f7d472dc1f

Test Train Split Architecture

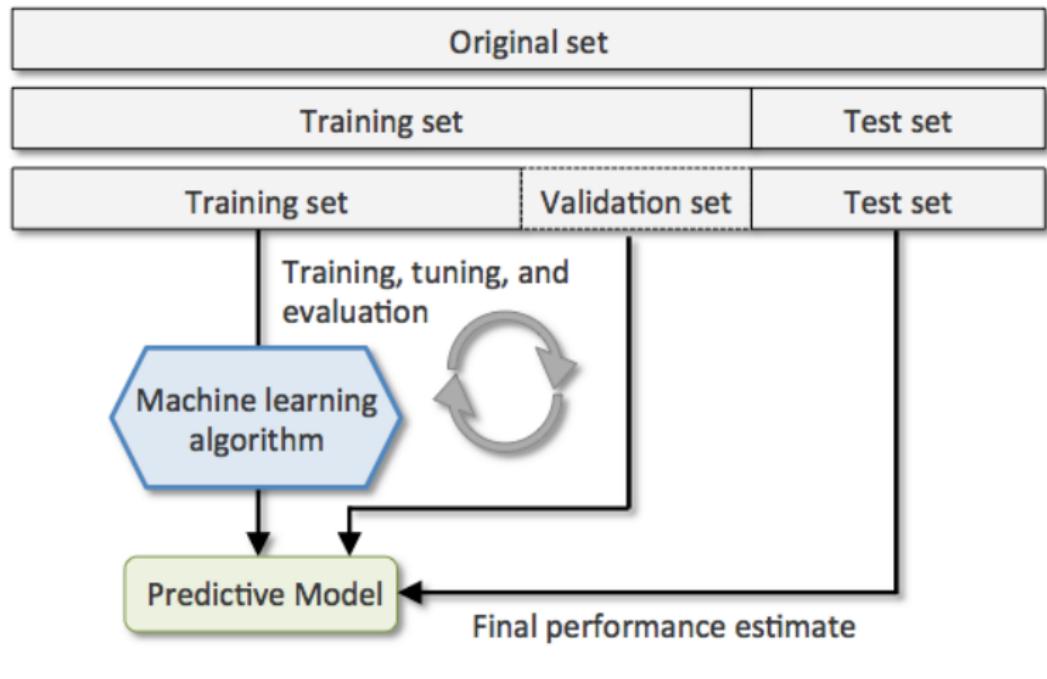


Fig 3: Test Train Arch. src :

medium.com/datadriveninvestor/data-science-essentials-why-train-validation-test-data-b7f7d472dc1f

Model Evaluation

Model Evaluation

1. Confusion Matrix

Model Evaluation

1. Confusion Matrix
2. F1 Score

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC
6. Log Loss

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC
6. Log Loss
7. Gini Coefficient

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC
6. Log Loss
7. Gini Coefficient
8. Concordant – Discordant Ratio

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC
6. Log Loss
7. Gini Coefficient
8. Concordant – Discordant Ratio
9. Root Mean Squared Error

Model Evaluation

1. Confusion Matrix
2. F1 Score
3. Gain and Lift Charts
4. Kolmogorov Smirnov Chart
5. AUC – ROC
6. Log Loss
7. Gini Coefficient
8. Concordant – Discordant Ratio
9. Root Mean Squared Error
10. etc.

src : <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>

Confusion Matrix Example

n = 165	Predicted:	
	No	Yes
Actual: No	50	10
Actual: Yes	5	100

Fig 4: Confusion Matrix src : www.geeksforgeeks.org/confusion-matrix-machine-learning/

ML Model Test DEMO

Trying to Evade Single Feature Detection.

How can we evade?

EVADE ML
DETECTION

How can we evade?

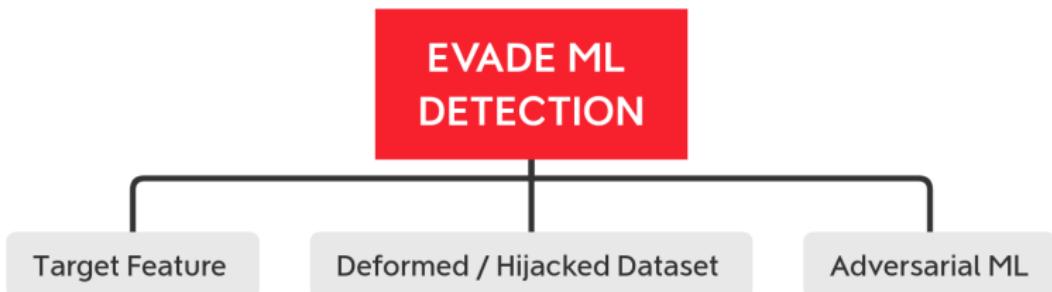
EVADE ML
DETECTION

Deformed / Hijacked Dataset

How can we evade?



How can we evade?



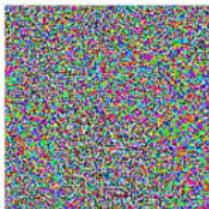
Generative Adversarial Networks



"panda"

57.7% confidence

$+ .007 \times$



"noise"

=



"raccoon"

99.3% confidence

src

: google images

GAN in Image Recognition



=



src : google images

Evasion PoC

(DevMode) x64mayhem\$python3 AdversarialTestDemo.py

ADVERSARIAL Malware Mutator - SAKET UPADHYAY

[[0 0 0 ... 0 0 0]]

The classification is = Malware

With accuracy of (Malware | Safe) = 94.0 | 6.0

(DevMode) x64mayhem\$python3 AdversarialTestDemo.py

ADVERSARIAL Malware Mutator - SAKET UPADHYAY

[[0 0 ... 0 0 0]]

The classification is = **Malware**

With accuracy of (Malware | Safe) = **94.0** | 6.0

ROUND [1] : Activated Bit Count = 101 setting random limit = 50

pass = 407 || Activated Bit Count = 151

ROUND [2] : Activated Bit Count = 101 setting random limit = 50

pass = 19732 || Activated Bit Count = 151

ROUND [3] : Activated Bit Count = 101 setting random limit = 50

pass = 77 || Activated Bit Count = 150

ROUND [4] : Activated Bit Count = 101 setting random limit = 50

pass = 286 || Activated Bit Count = 151

ROUND [5] : Activated Bit Count = 101 setting random limit = 50

pass = 7636 || Activated Bit Count = 151

MUTATION COMPLETED

=====

best mutation = 150

The classification is = **Safe**

With accuracy of (Malware | Safe) = 26.1 | **73.9**

P.A.C.E.

Platform for Android Malware Classification and Performance Evaluation

Platform for Android Malware Classification and Performance Evaluation

PACE, a unified solution to offer **open and easy implementation access to several machine learning-based Android malware detection techniques** that make most of the research in this domain reproducible. The benefits of PACE are offered using three interfaces i.e. through REST API, Web Interface and ADB interface.

Platform for Android Malware Classification and Performance Evaluation

PACE, a unified solution to offer **open and easy implementation access to several machine learning-based Android malware detection techniques** that make most of the research in this domain reproducible. The benefits of PACE are offered using three interfaces i.e. through REST API, Web Interface and ADB interface.

Presented in: *The 3rd International Workshop on Big Data Analytic for Cyber Crime Investigation and Prevention, IEEE International Conference on Big Data 2019*

P.A.C.E.

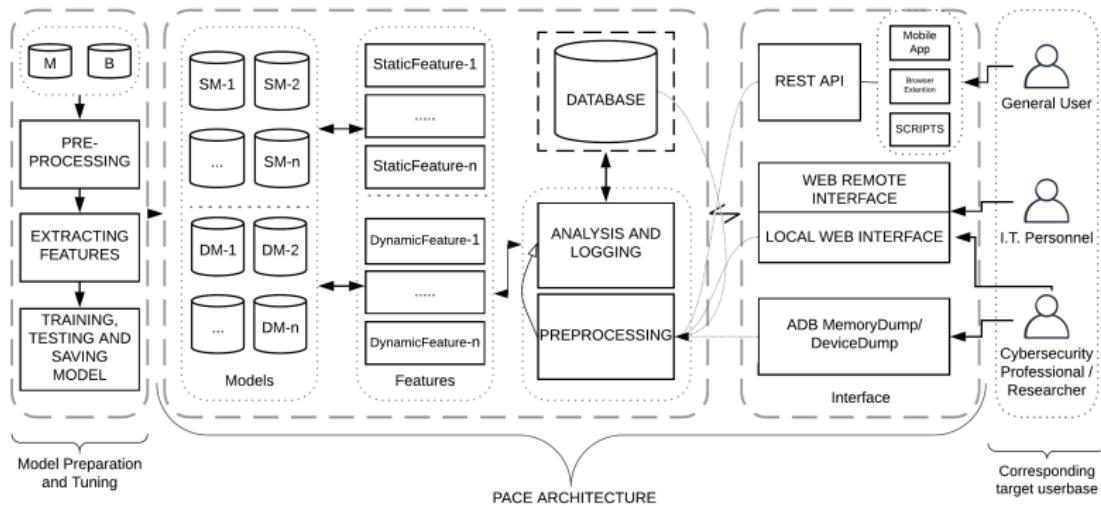


Fig 5: P.A.C.E. Architecture

PACE

UPLOAD FOR ANALYSIS

Browse...

No file selected.

Upload and analyse

Dashboard

Dashboard / Overview

ChartsTables

20

Permissions



Download Source Code

Download



0

SAFE HITS

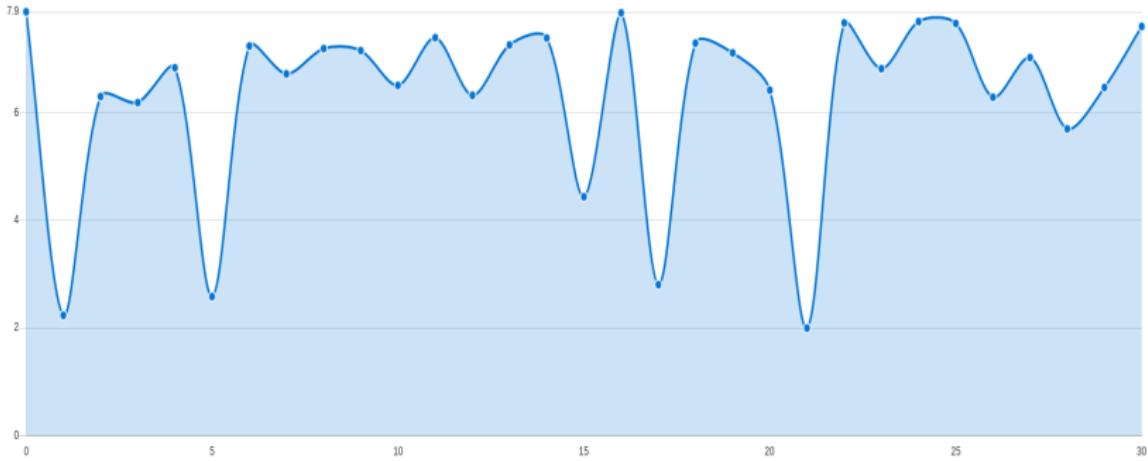


8

MALWARE HITS



Entropy Chart SHA256: dd3b7f7b15f22e249bcf18538082d4b20c763914cd7f29209ef8629e826d661



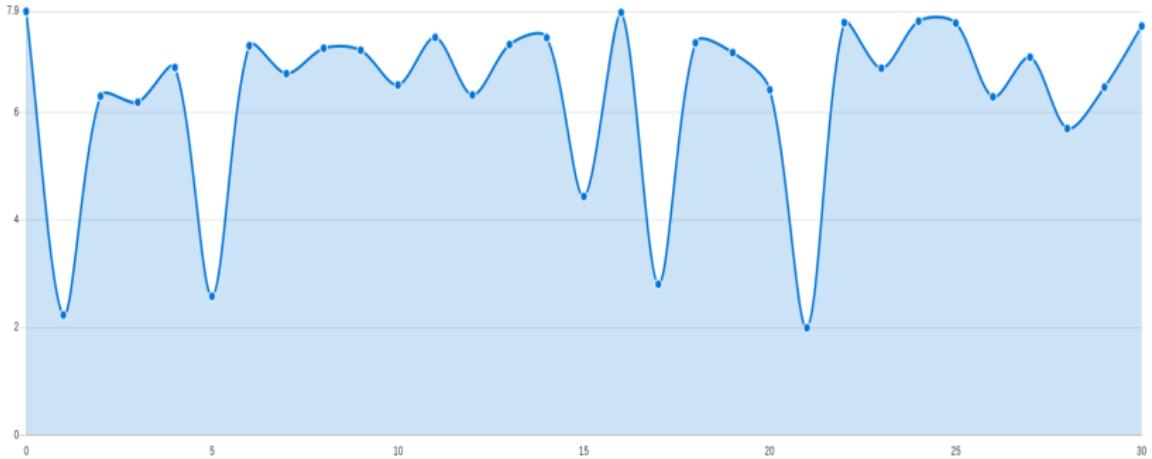
Updated 23/10/2019 22:57:16

ML Engine Results

No.	Name	Classification [Confidence]	Feature Type	Base Operation	Author

ML Engine Results

No.	Name	Classification [Confidence]	Feature Type	Base Operation	Author
1	Random Forest (88%)	MALWARE [67.60000000000001]	Permissions	Random Forest	Bhawna Yadav
2	SVM (83%)	MALWARE [92.80000000000001]	Permissions	Support Vector Machine	Saket Upadhyay
3	KNN (78%)	MALWARE [80.0]	Permissions	K-Nearest Neighbours	Wang, Wei and Li, Yuanyuan and Wang, Xing and Liu, Jiqiang and Zhang, Xiangliang
4	BAGGING (90%)	MALWARE [98.8]	Permissions	DT	Naser Peiravian and Xingquan Zhu
5	PEPB (84%)	MALWARE [98.8]	Permissions	Support Vector Machine	Chun-Ying Huang, Yi-Ting Tsai, and Chung-Han Hsu
6	ADABOOST-GRADBOOST (84%)	MALWARE [98.8]	Permissions	Gradient Boosting	Yerima, Suleiman Y and Sezer, Sakir
7	APK Auditor: Permission-based Android malware detection (79%)	MALWARE [94.39999999999999]	Permissions	Logistic regression	----
8	Permission-Based Android Malware Detection (89%)	NA []	Permissions	K-means(Random forest)	Zarni Aung, Win Zaw
9	STATIC DETECTION OF ANDROID MALWARE BY USING PERMISSIONS AND API CALLS (88%)	MALWARE [69.19999999999999]	Permissions	Random Forest	Patrick P.K. Chan, Wen-Kai Song
No.	Name	Classification [Confidence]	Feature Type	Base Operation	Author



*Representational file entropy

Bar Chart

15000

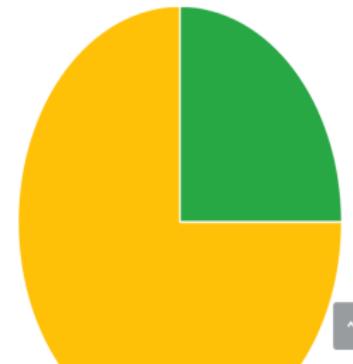
10000

5000



Confidence Distribution

SAFE SUSPICIOUS MALWARE



Dashboard

Dashboard / Tables

Charts

Tables

Permissions

No.	Name	Stock	okay?	Normal?	Title
1	android.permission.VIBRATE	True	--	--	--
2	android.permission.WRITE_EXTERNAL_STORAGE	False	--	--	--
3	android.permission.WAKE_LOCK	True	--	--	--
4	android.permission.CAMERA	False	--	--	--
5	android.permission.READ_PHONE_STATE	False	--	--	--
6	android.permission.FOREGROUND_SERVICE	False	--	--	--
7	android.permission.REQUEST_DELETE_PACKAGES	False	--	--	--
8	android.permission.MOUNT_UNMOUNT_FILESYSTEMS	True	--	--	--
9	android.permission.ACCESS_NETWORK_STATE	True	--	--	--
10	android.permission.RECEIVE_BOOT_COMPLETED	True	--	--	--
11	android.permission.BIND_ACCESSIBILITY_SERVICE	True	--	--	--
12	com.android.launcher.permission.INSTALL_SHORTCUT	False	--	--	--
13	android.permission.GET_PACKAGE_SIZE	True	--	--	--
14	android.permission.REQUEST_INSTALL_PACKAGES	True	--	--	--
15	android.permission.FLASHLIGHT	True	--	--	--
16	android.permission.ACCESS_WIFI_STATE	True	--	--	--
17	com.google.android.c2dm.permission.RECEIVE	False	--	--	--

Exif Info

No.	Property	Value	--	--	--
1	File Name	APKPure_v3.12.1_apkpure.com.apk	--	--	--
2	File Size	12 MB	--	--	--
3	File Modification Date/Time	2019	--	--	--
4	File Access Date/Time	2019	--	--	--
5	File Inode Change Date/Time	2019	--	--	--
6	File Permissions	rW-r-r-	--	--	--
7	File Type	ZIP	--	--	--
8	File Type Extension	zip	--	--	--
9	MIME Type	application/zip	--	--	--
10	Zip Required Version	20	--	--	--
11	Zip Bit Flag	0x0808	--	--	--
12	Zip Compression	Deflated	--	--	--
13	Zip Modify Date	2019	--	--	--
14	Zip CRC	0x8f17ef4a	--	--	--
15	Zip Compressed Size	77719	--	--	--
16	Zip Uncompressed Size	225152	--	--	--
17	Zip File Name	META-INF/MANIFEST.MF	--	--	--
No.	Property	Value	--	--	--

Updated yesterday at 11:59 PM



Dashboard

Dashboard / Overview

Charts

Tables

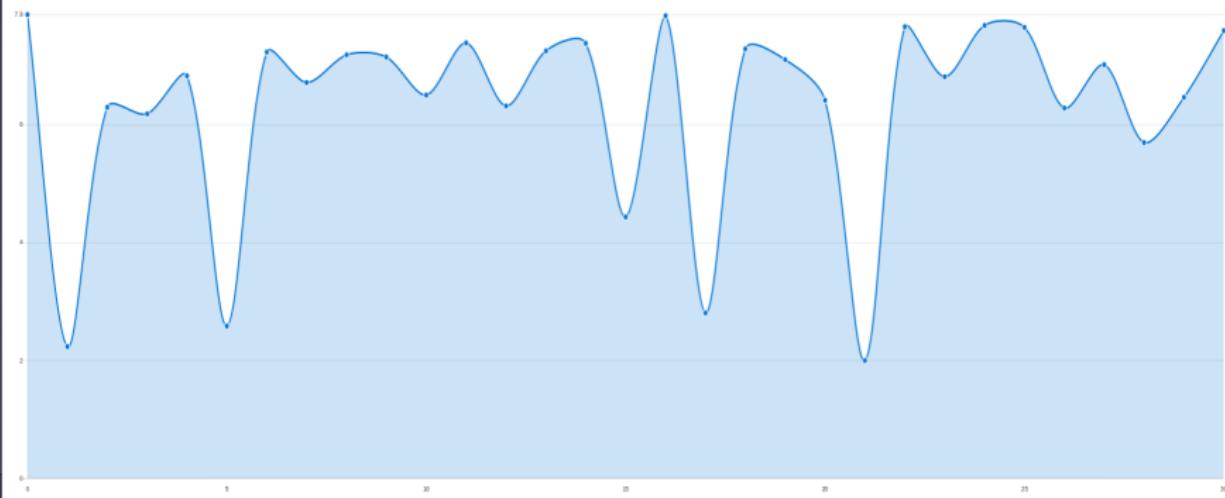
20
Permissions

Download Source Code
SHA256: dd3b7f7b15922e2408cf1853808034b20c763914d725209e8826e826661
Signed

0
SAFE HITS

8
MALWARE HITS

Entropy Chart SHA256: dd3b7f7b15922e2408cf1853808034b20c763914d725209e8826e826661



Updated 23/10/2019 22:57:18

ML Engine Results

No.	Name	Classification [Confidence]	Feature Type	Base Operation	Author
1	Random Forest (88%)	MALWARE [67.60000000000001]	Permissions	Random Forest	Bhawna Yadav
2	SVM (83%)	MALWARE [92.80000000000001]	Permissions	Support Vector Machine	Saleet Upadhyay
3	KNN (78%)	MALWARE [98.0]	Permissions	K-Nearest Neighbours	Wang, Wei and Li, Yuanxuan and Wang, Xing and Liu, Jiepong and Zhang, Xianglong
4	BAGGING (90%)	MALWARE [98.8]	Permissions	DT	Naser Perwani and Xingshan Zhu
5	PEPB (84%)	MALWARE [98.8]	Permissions	Support Vector Machine	Chun-Ying Huang, Yi-Ting Tsai, and Chung-Han Hsu
6	ADABOOST-GRADBOOST (84%)	MALWARE [98.8]	Permissions	Gradient Boosting	Yelima, Suleiman Y and Sezer, Salir

Conclusion

Closing Thoughts

Closing Thoughts

- Malware Detection and Elimination plays a critical role in defending critical digital infrastructure.

Closing Thoughts

- Malware Detection and Elimination plays an critical role in defending critical digital infratucture.
- With advantage of AI & ML,Cyber- and Info -sec field is adapting significantly.

Closing Thoughts

- Malware Detection and Elimination plays an critical role in defending critical digital infratucture.
- With advantage of AI & ML,Cyber- and Info -sec field is adapting significantly.
- The defenders as well as adversaries have access to new tech. and research.

Closing Thoughts

- Malware Detection and Elimination plays an critical role in defending critical digital infratucture.
- With advantage of AI & ML,Cyber- and Info -sec field is adapting significantly.
- The defenders as well as adversaries have access to new tech. and research.
- There's lot more to explore.

Due Credits and Mentions

Due Credits and Mentions

Co-Authors in the above discussed research.(PACE)

Dr. Ajit Kumar

Dr. Shishir K. Shandilya

Dr. Andrii Shalaginov

Dr. Vinti Agarwal

Bhawna Yadav

Thankyou

Q & A



x64mayhem



linkedin.com/in/saketupadhyay



x64mayhem@gmail.com



github.com/Saket-Upadhyay

Resources at : github.com/Saket-Upadhyay/PennStateTalk2020

Q & A



x64mayhem



linkedin.com/in/saketupadhyay



x64mayhem@gmail.com



github.com/Saket-Upadhyay

Resources at : github.com/Saket-Upadhyay/PennStateTalk2020