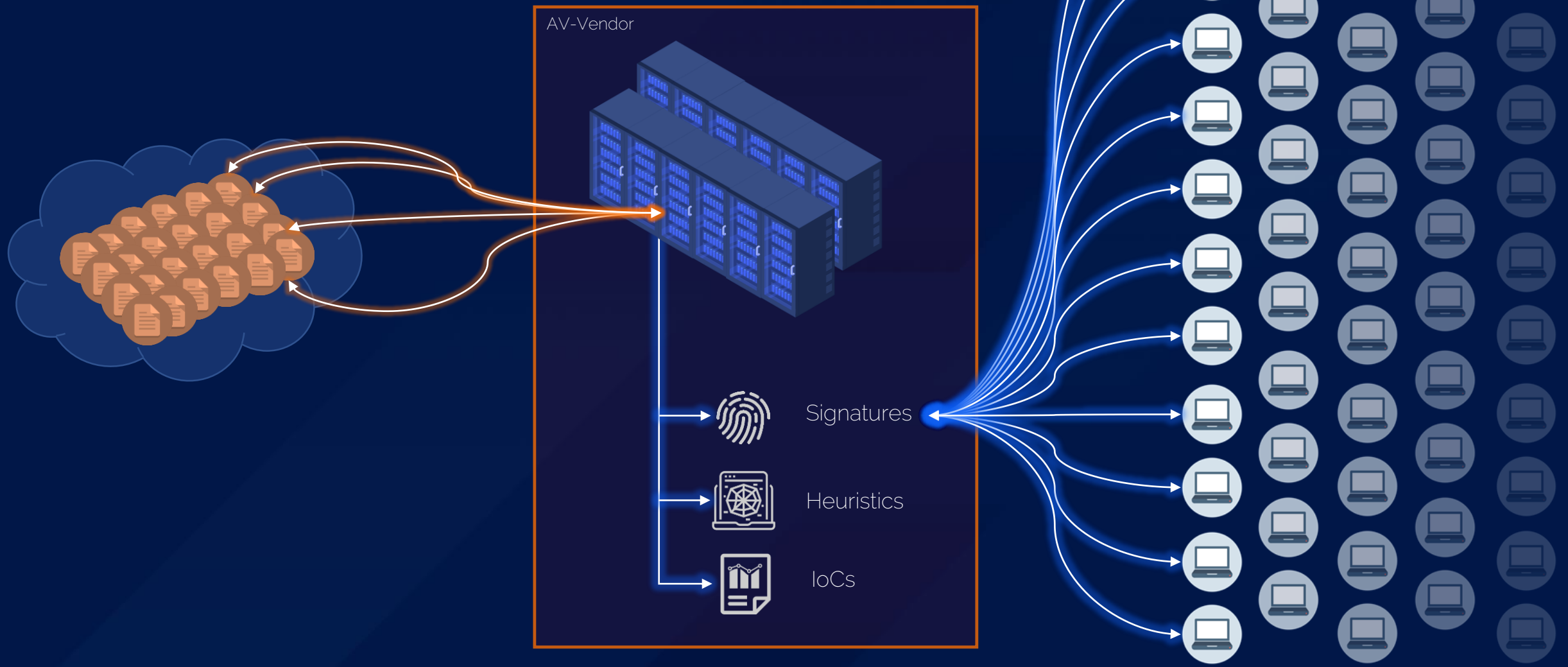
MALWAREERKENNUNG VIA COMPUTERVISION

ANTIVIRUS
HOW DOES IT WORK



CURRENT SITUATION

A FUNDAMENTAL CHANGE IN HOW CYBER ATTACKS CONDUCTED

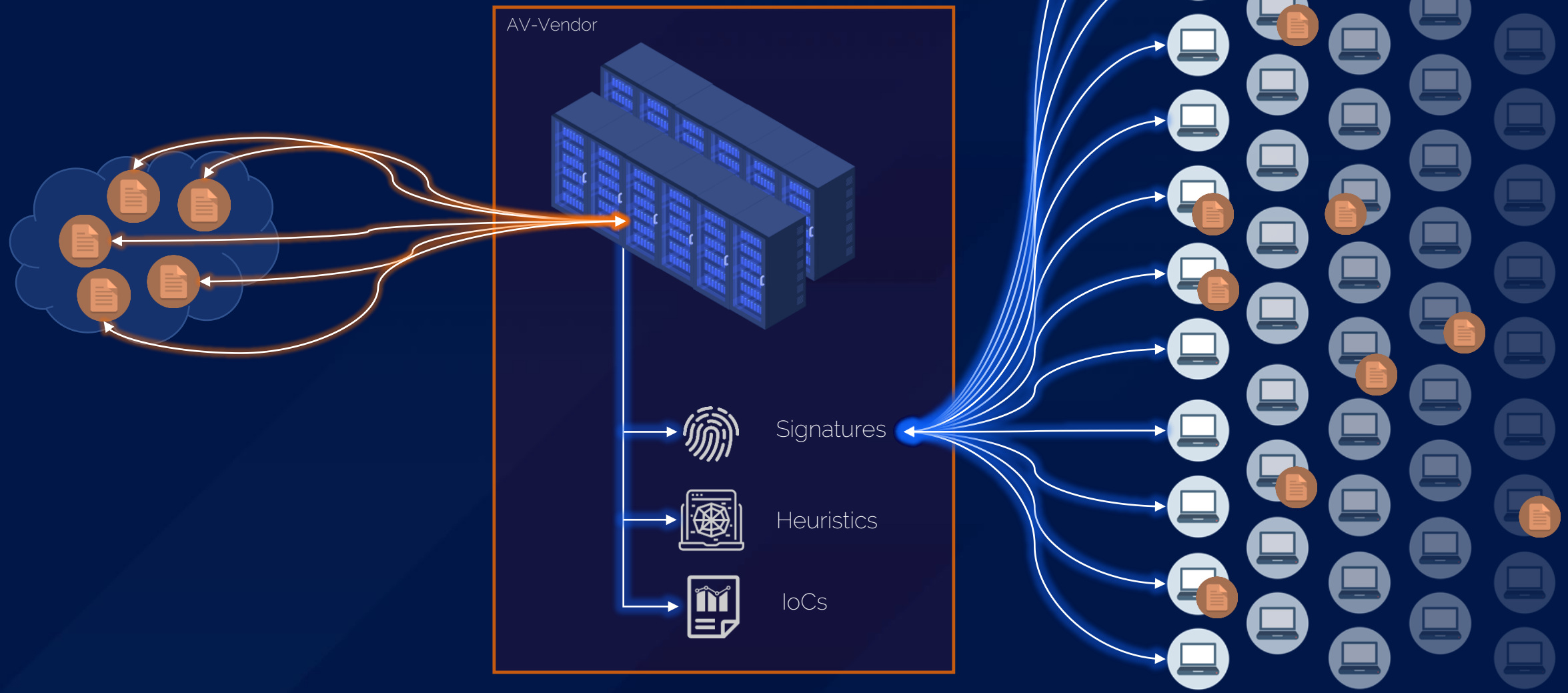
360K "NEW MALWARE SAMPLES HIT THE SCENE EVERY DAY"
KASPERSKY

70% "OF MALWARE ONLY EXISTS ONCE"
FIREEYE

82% "OF MALWARE DISAPPEAR AFTER ONE HOUR"
KASPERSKY

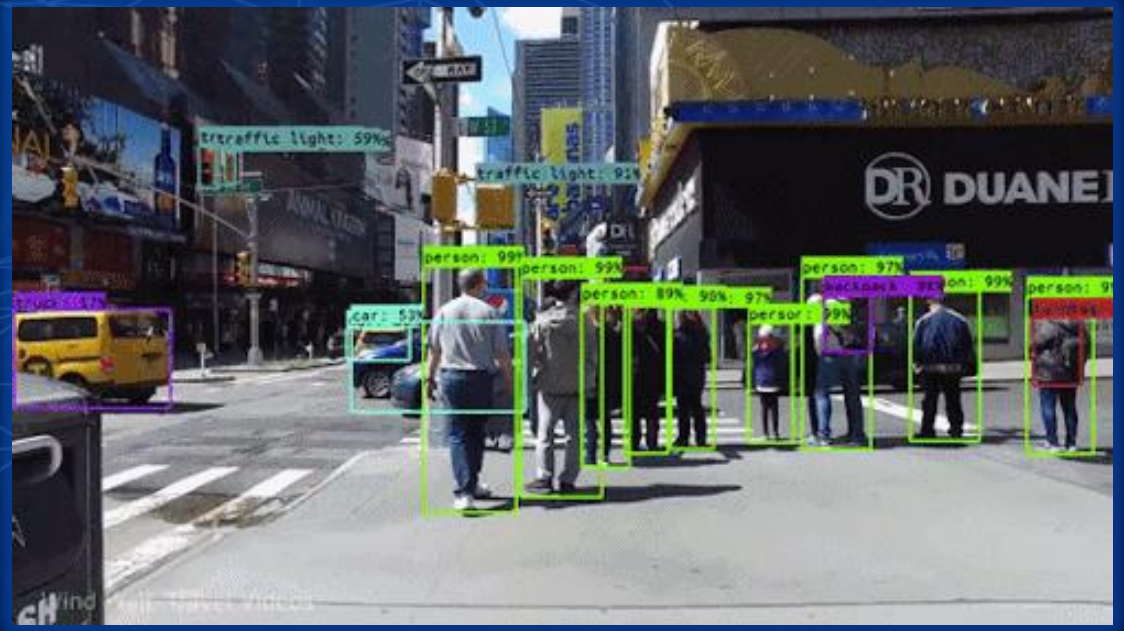
INLYSE

ANTIVIRUS
HOW DOES IT WORK



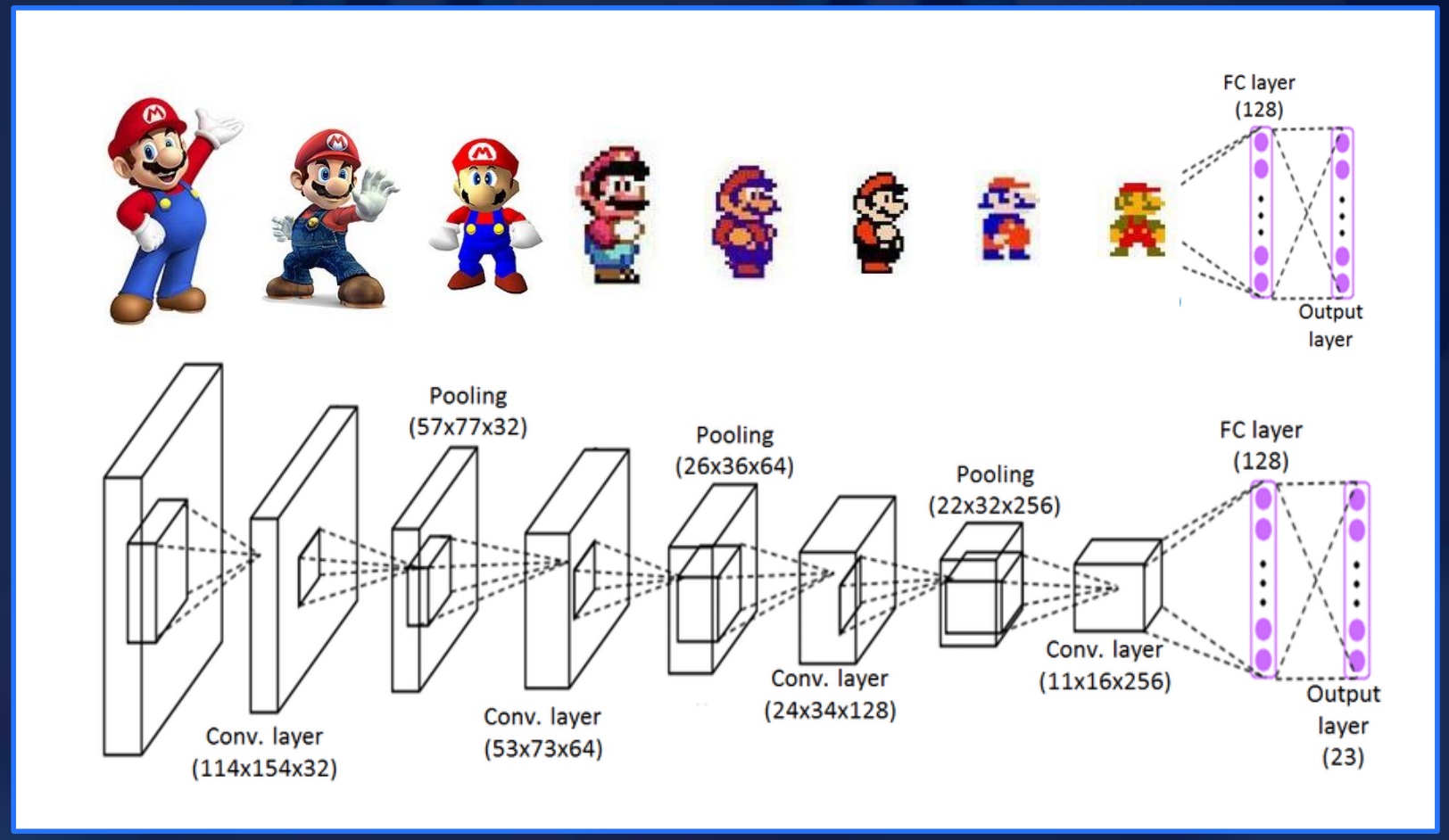
COMPUTER VISION

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs.



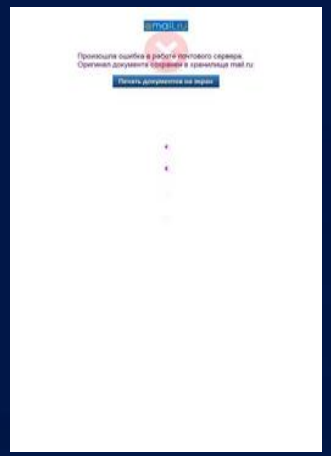
INLYSE

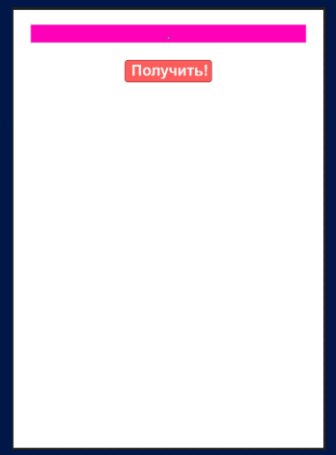
CNN
CONCEPT

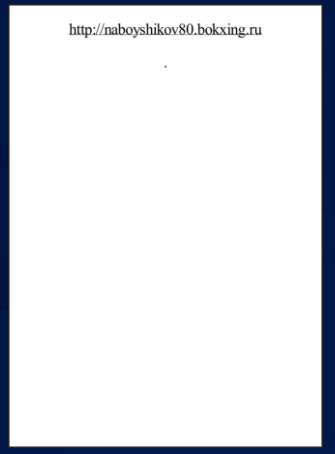


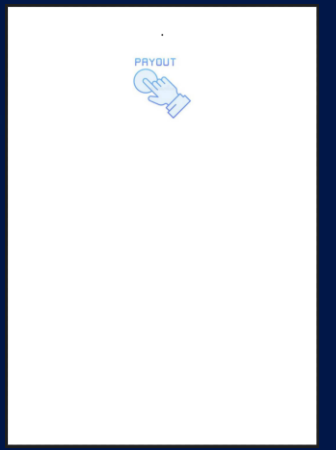
INLYSE

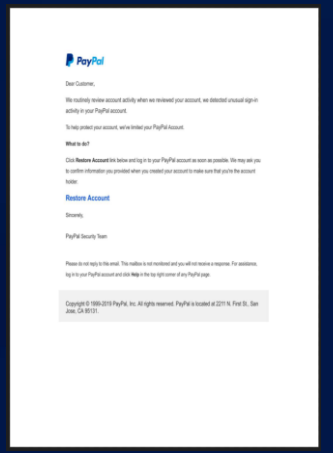
EXAMPLES
PHISHING/SCAM

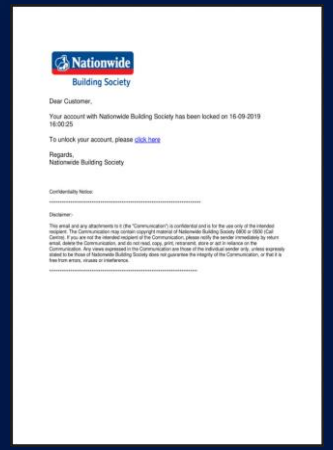












CNN
PHISHING

Setup

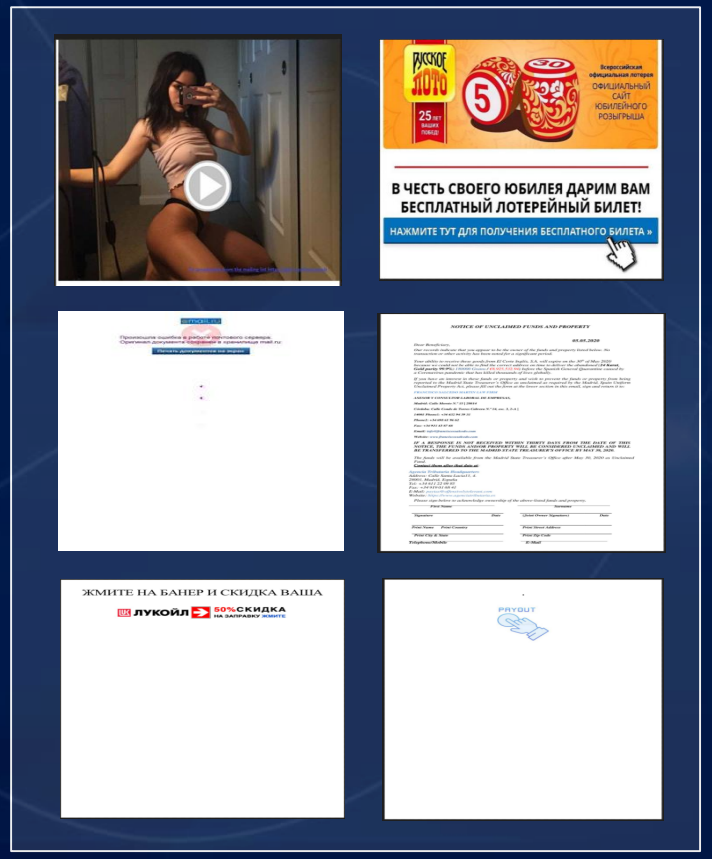
1. Data Selection

   50k unique Phishing PDFs50k unique benign PDFs

2. Convert first page into jpg
3. Resize images to 256x256 pixels
4. Train a DL Model

   → *Train-Test-Split = 80:20*

5. Validate



INLYSE

CNN PHISHING

Setup

1. Data Selection
   50k unique Phishing PDFs50k unique benign PDFs
2. Convert first page into jpg
3. Resize images to 256x256 pixels
4. Train a DL Model
   → *Train-Test-Split = 80:20*
5. Validate

Results

92%

Correct classified

INLYSE

DATA VISUALIZATION

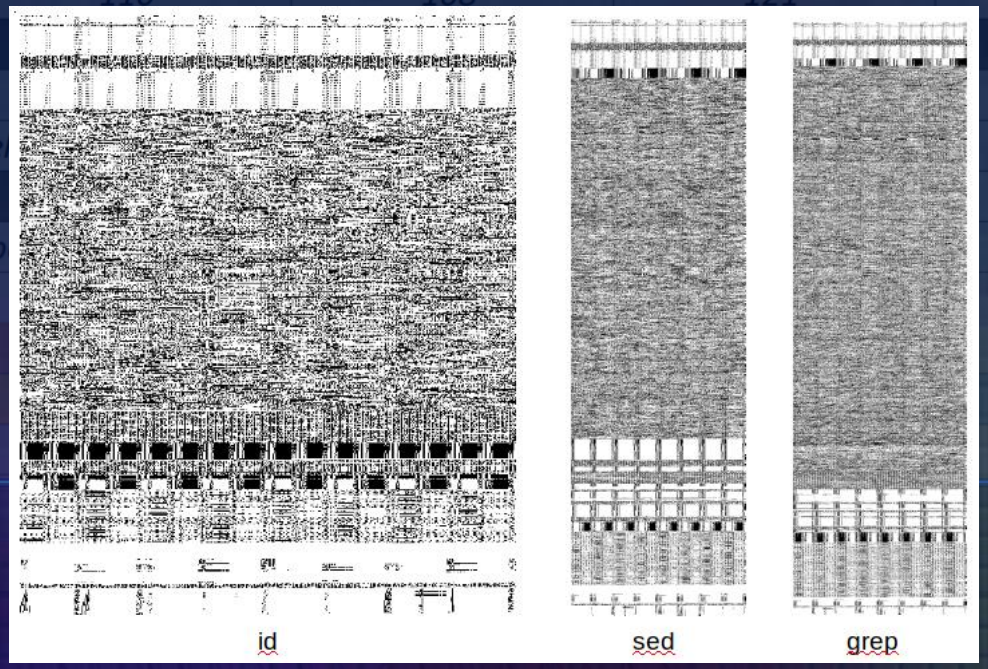
ASCII	i								n								l								y								s								e							
Binary	0	1	1	0	1	0	0	1	0	1	1	0	1	1	1	0	0	1	1	0	1	1	0	0	0	1	1	1	1	0	0	1	0	1	1	1	0	0	1	1	0	1	1	0	0	1	0	1
Bit by Bit																																																	
Decimal	*105*								*110*								*108*								*121*								*115*								*101*							
GrayScale																																																
inlyse	*Multiple Conversions.*																																															
Allobyte	*Byte Tuples to RGB Pixels*																																															
Allobyte RGB	*(105, 110, 108)*								*(121, 115, 101)*																																									
Evosense	*Generates Color Table based on the distance of the bytes of input data*																																															
Evosense	*a b c d e f g h i j k l m n o p q r s t u v w x y z*																																															
Color Map																																																
Evosense RGB	*i*								*n*								*l*								*y*								*s*								*e*							
Evosense RGB	i, n, l								y, s, e																																								

Evosense RGB i, n, l y, s, e
Evosense RGB
Color Map
Evosense

INLYSE

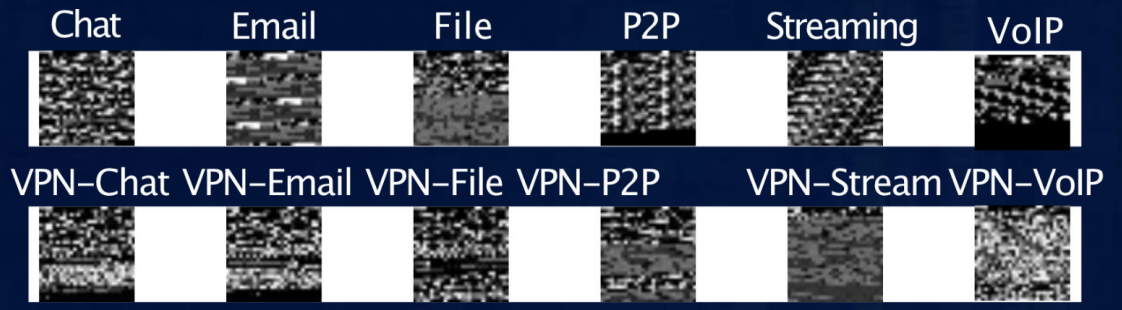
DATA VISUALIZATION

ASCII	i								n								l								y								s								e							
Binary	0	1	1	0	1	0	0	1	0	1	1	0	1	1	1	0	0	1	1	0	1	1	0	0	0	1	1	1	1	0	0	1	0	1	1	1	0	0	1	1	0	1	1	0	0	1	0	1
Bit by Bit																																																	
Decimal	*105*								*110*								*108*								*121*								*115*								*101*							
GrayScale																																																
inlyse	*Multiple Conversions.*																																															
Allobyte	*Byte Tuples to RGB Pixe*																																															
Allobyte RGB	*(105, 110, 108)*																																															
Evosense	*Generates Color Table b*																																															
Evosense	*a b c d e f g h i*																																															
Color Map																																																
Evosense RGB	*i*																																															
Evosense RGB	i, n, l																																															

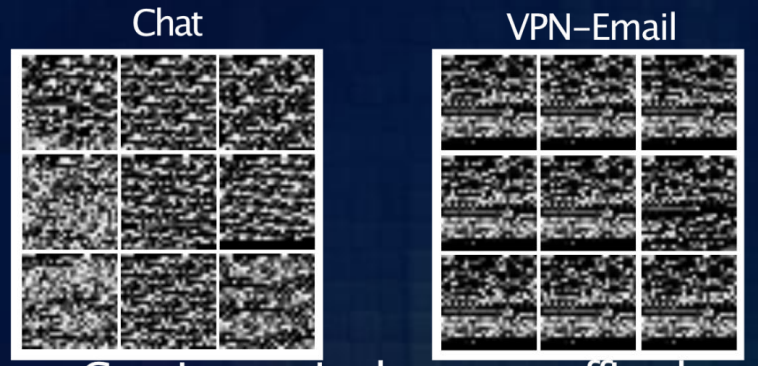


INLYSE

DATA CLASSIFICATION



Visualization of all classes of traffic

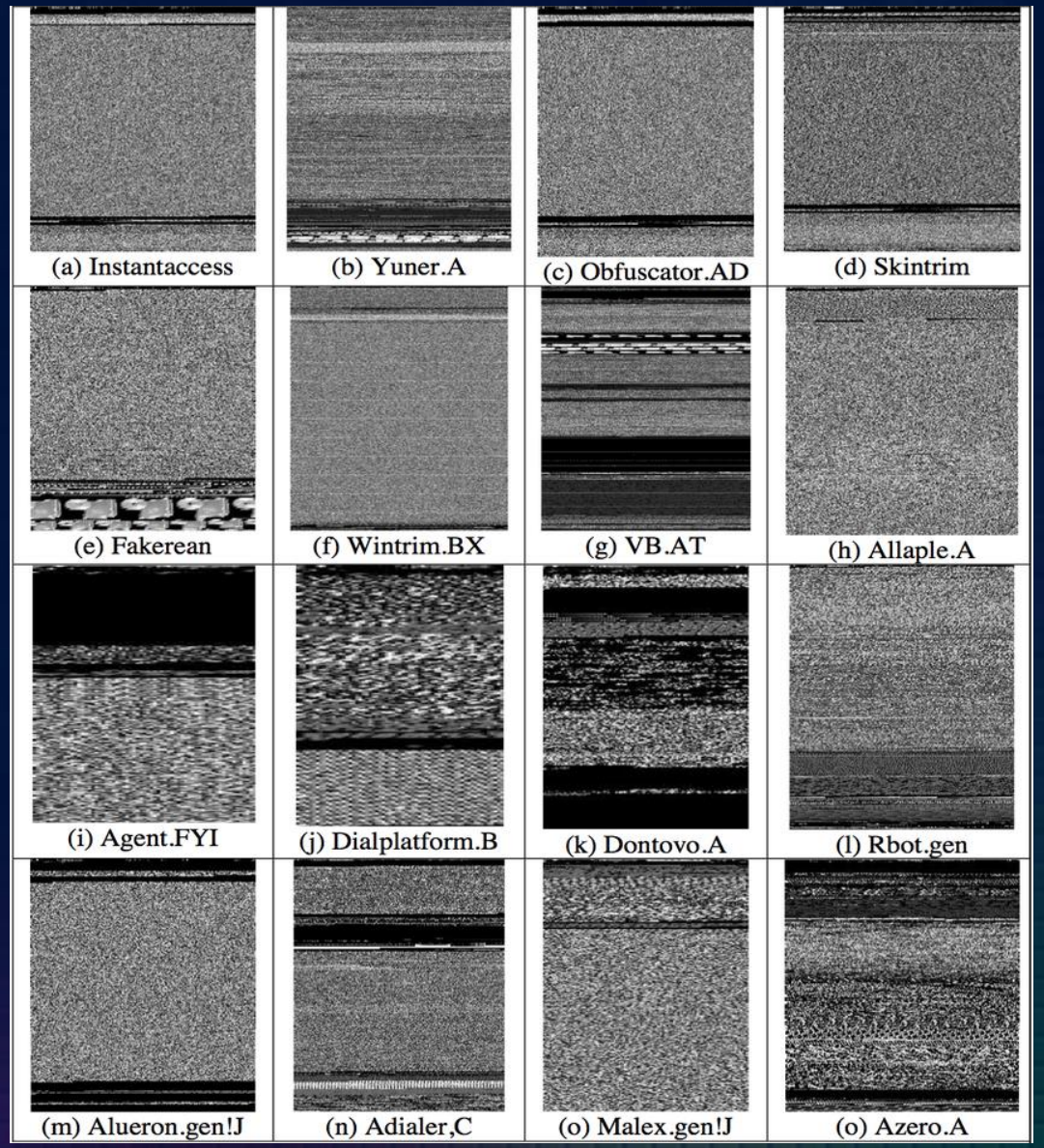


Consistency in the same traffic class

Visualization of Encrypted Traffic

INLYSE

MALWARE CLASSIFICATION

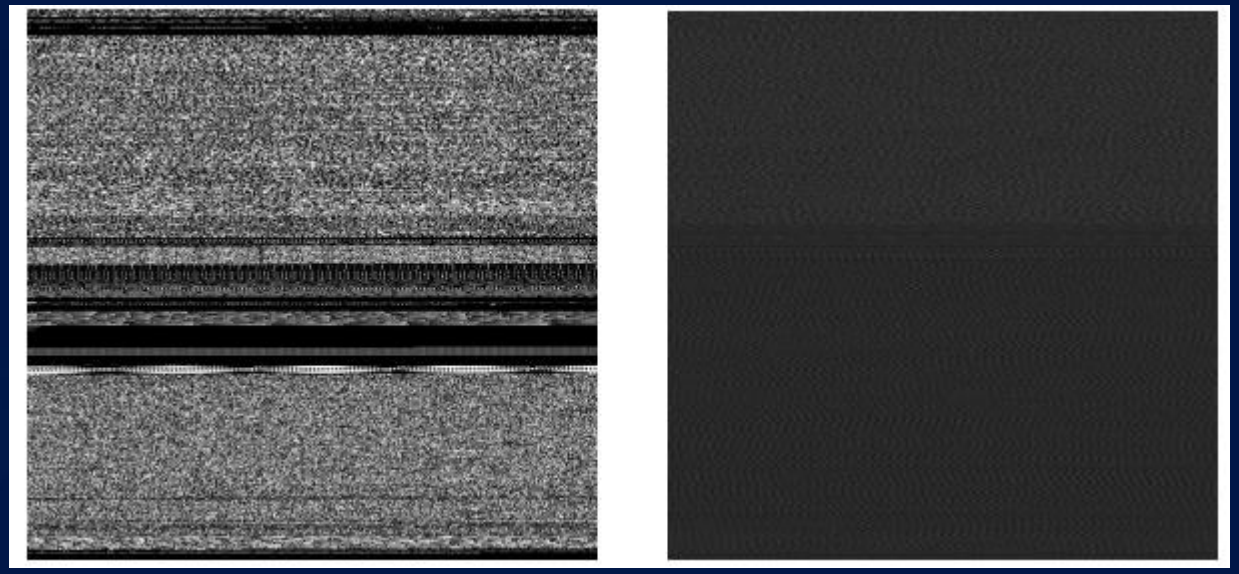


INLYSE

KAGGLE
Microsoft

Challenge

Classification of 500GB of malware into 9 malware family classes.

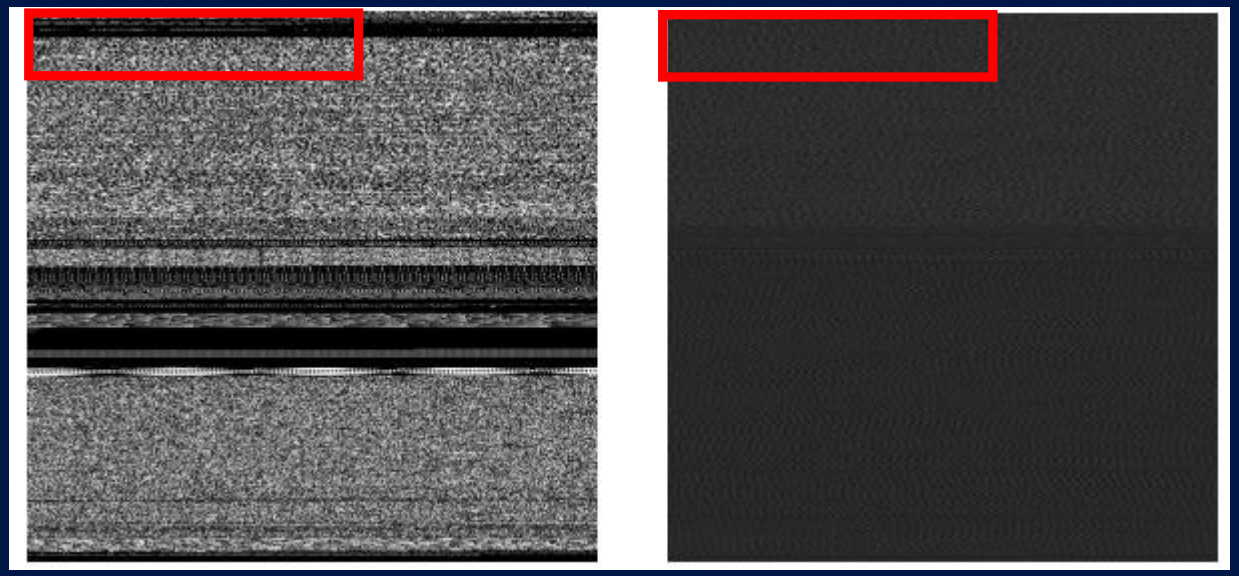


INLYSE

KAGGLE
Microsoft

Challenge

Classification of 500GB of malware into 9 malware family classes.



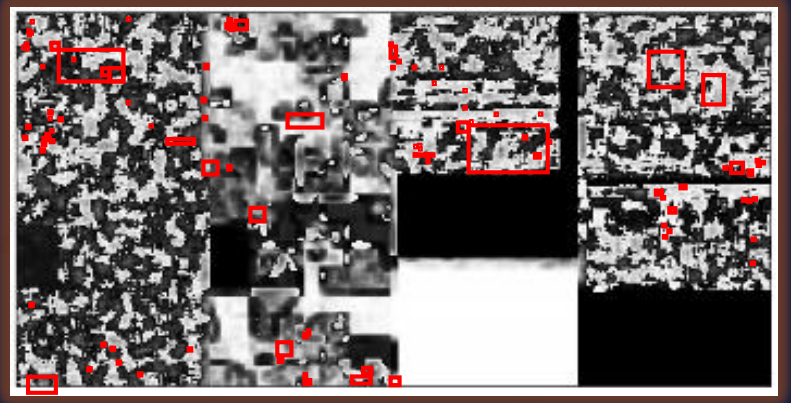
Winning Solution:

>99% correct classified

INLYSE

MALWARE IN REAL LIFE DATA

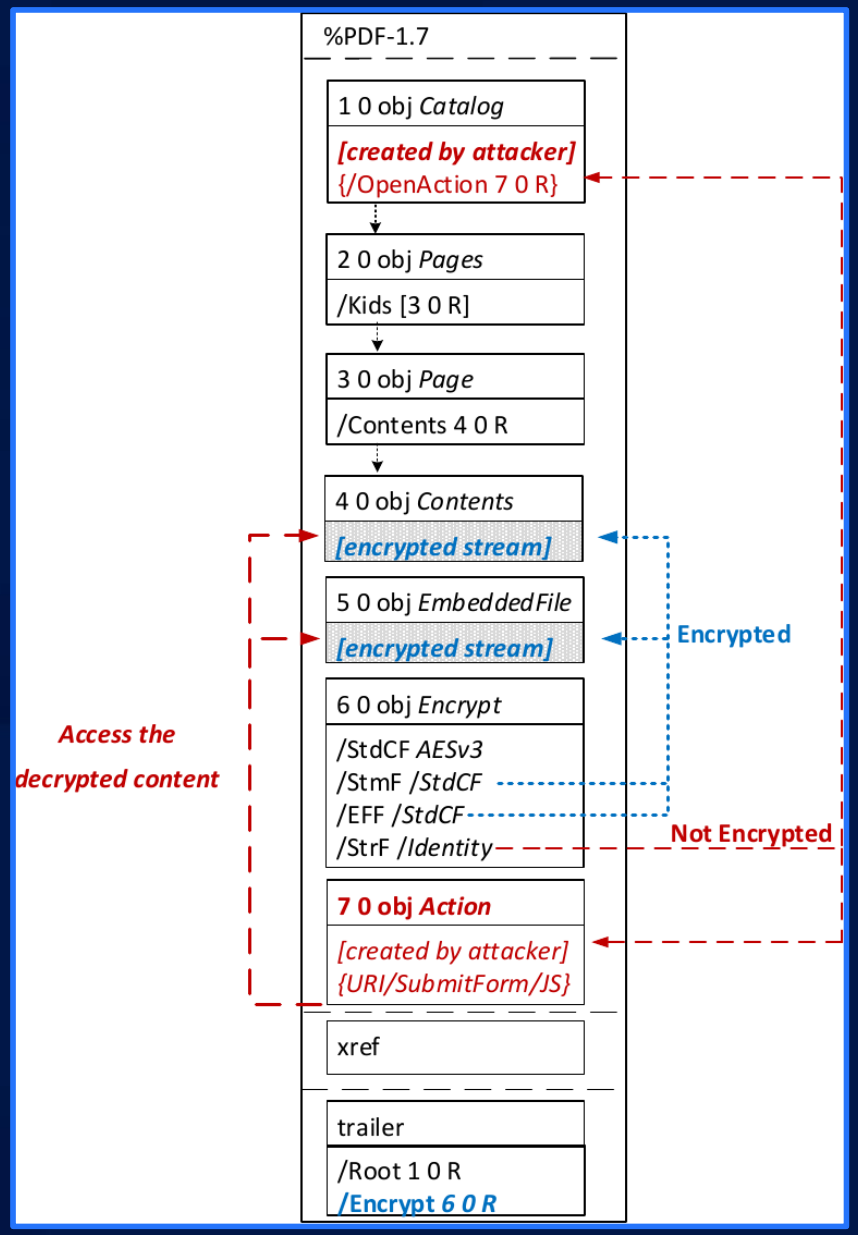
Hides itself as much as possible within data to circumvent its detection.



DATA AND FILES

Multitude of different encodings and filters

- Encodings: ASCII, Hex, Base64,
- Filters: Compression, Encryption,
- More Datatypes: Images, Videos, Attachments.



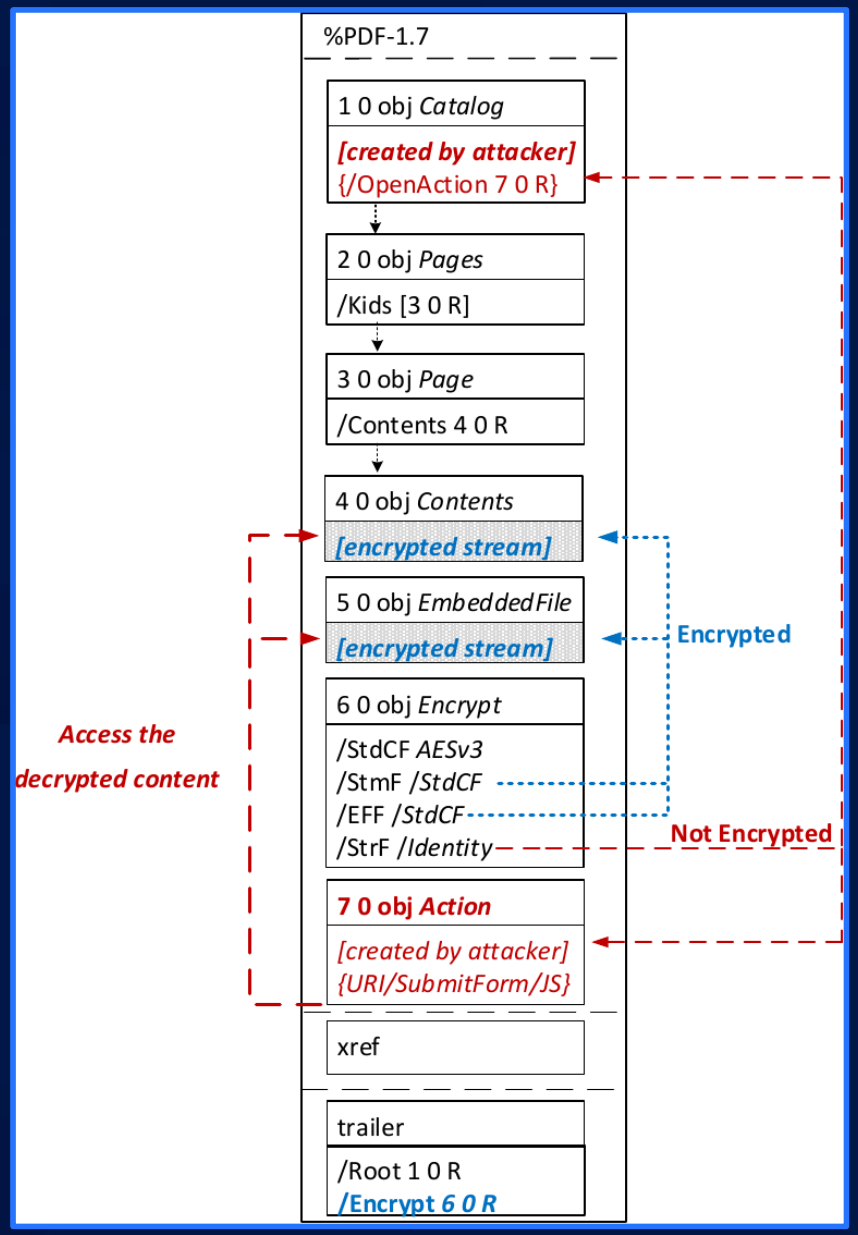
INLYSE

DATA AND FILES

Multitude of different encodings and filters

- Encodings: ASCII, Hex, Base64,
- Filters: Compression, Encryption,
- More Datatypes: Images, Videos, Attachments.

Interpretation of image data -> context required

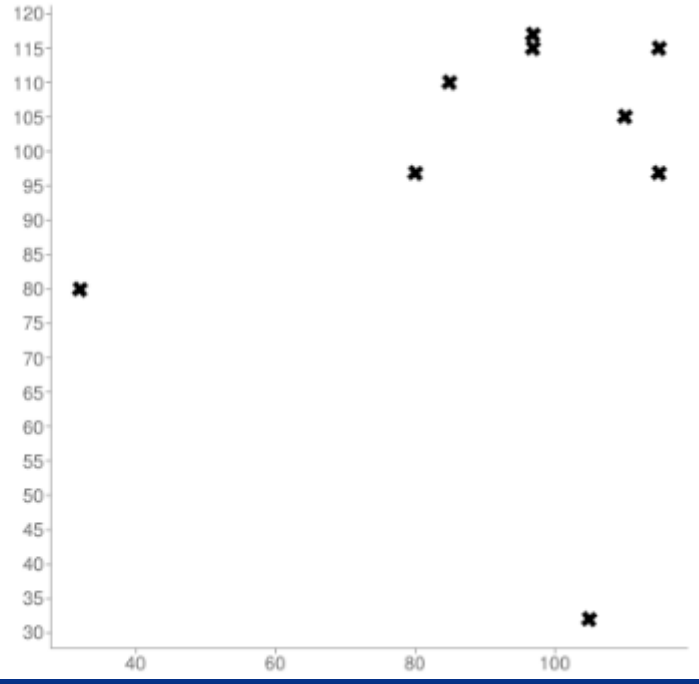


INLYSE

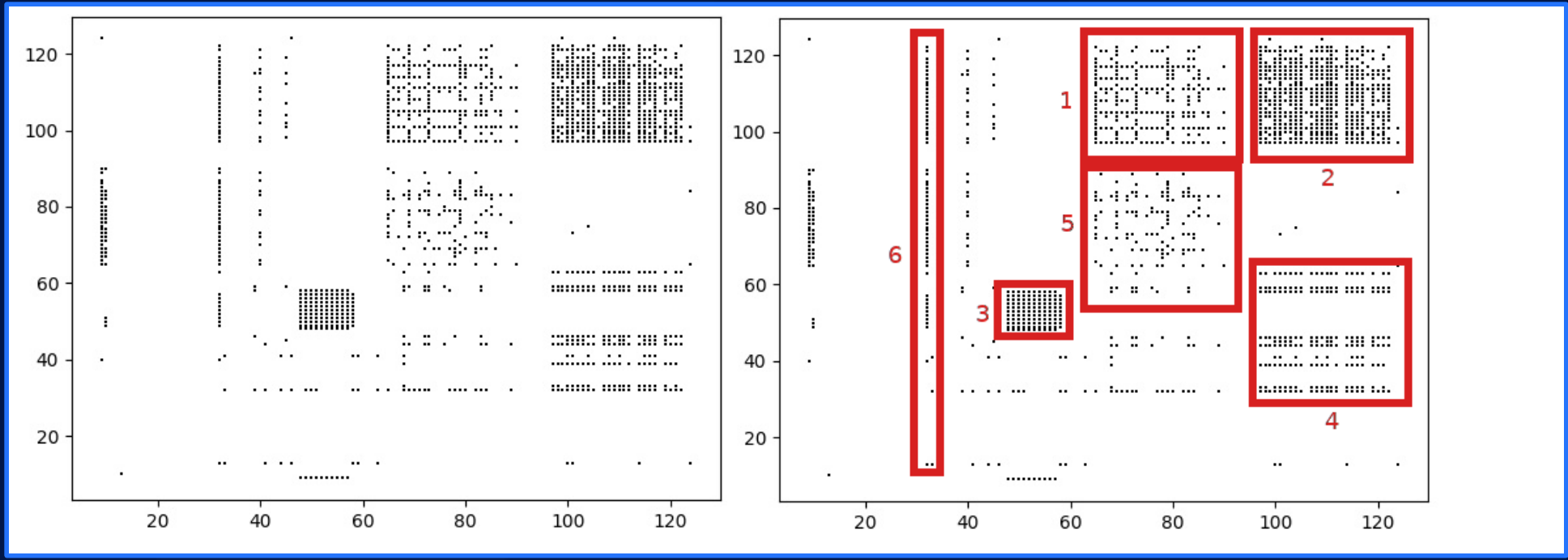
DATA CONTEXT

Context through statistical analysis.

Uni Passau	(x , y)
Un	(85 ,110)
ni	(110,105)
i	(105, 32)
P	(32 , 80)
Pa	(80 , 97)
as	(97 ,115)
ss	(115,115)
sa	(115, 97)
au	(97, 117)



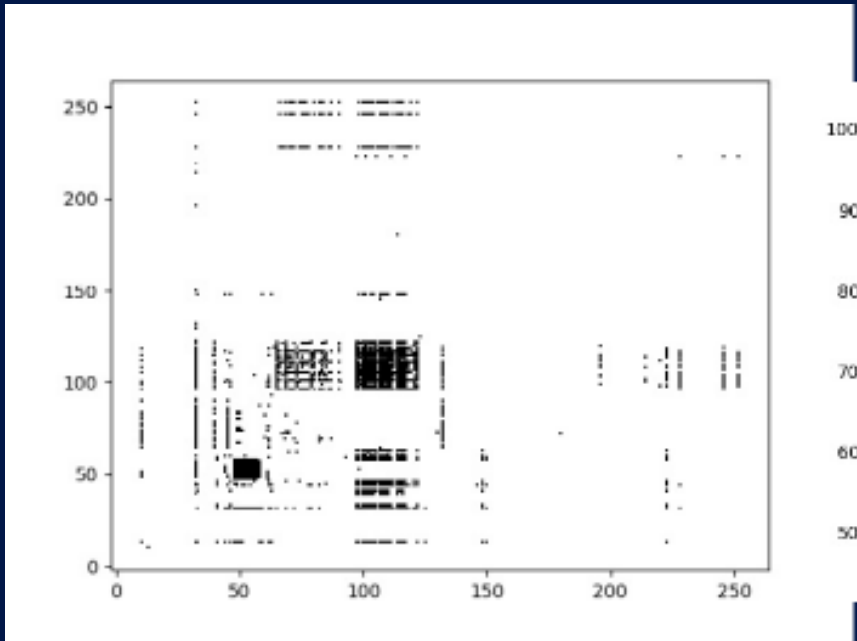
DATA CONTEXT



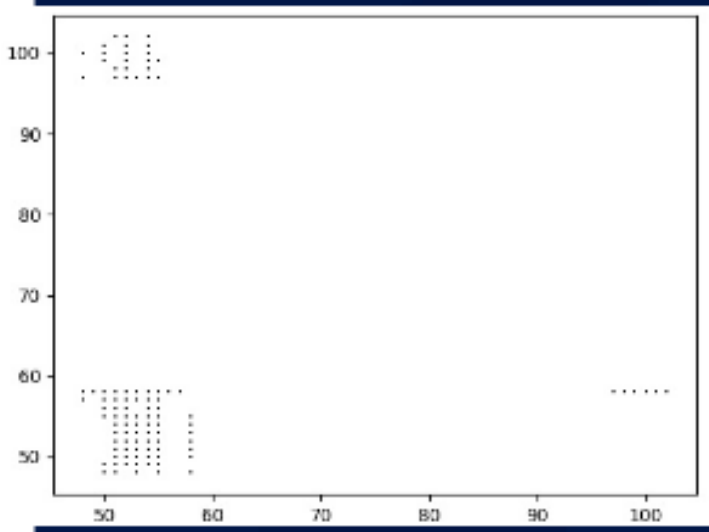
1. Upper + Lower Letter
2. Lower Letter + Lower Letter
3. Numbers + X
4. Special + Lower Letter
5. Special + Upper Letter
6. Spaces + Letters

DATA CONTEXT

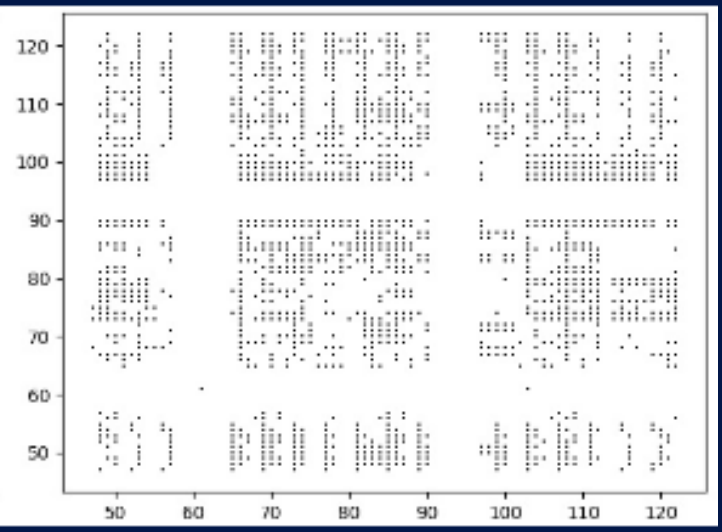
ASCII Text



HEX



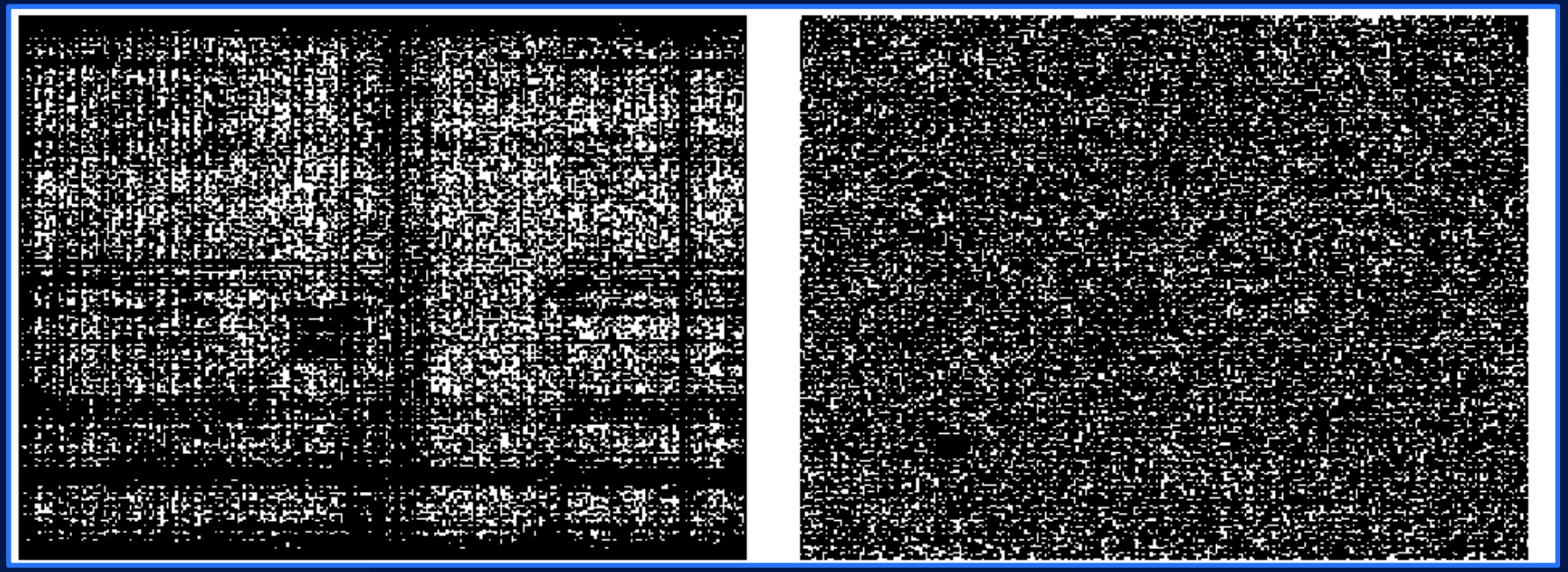
Base64



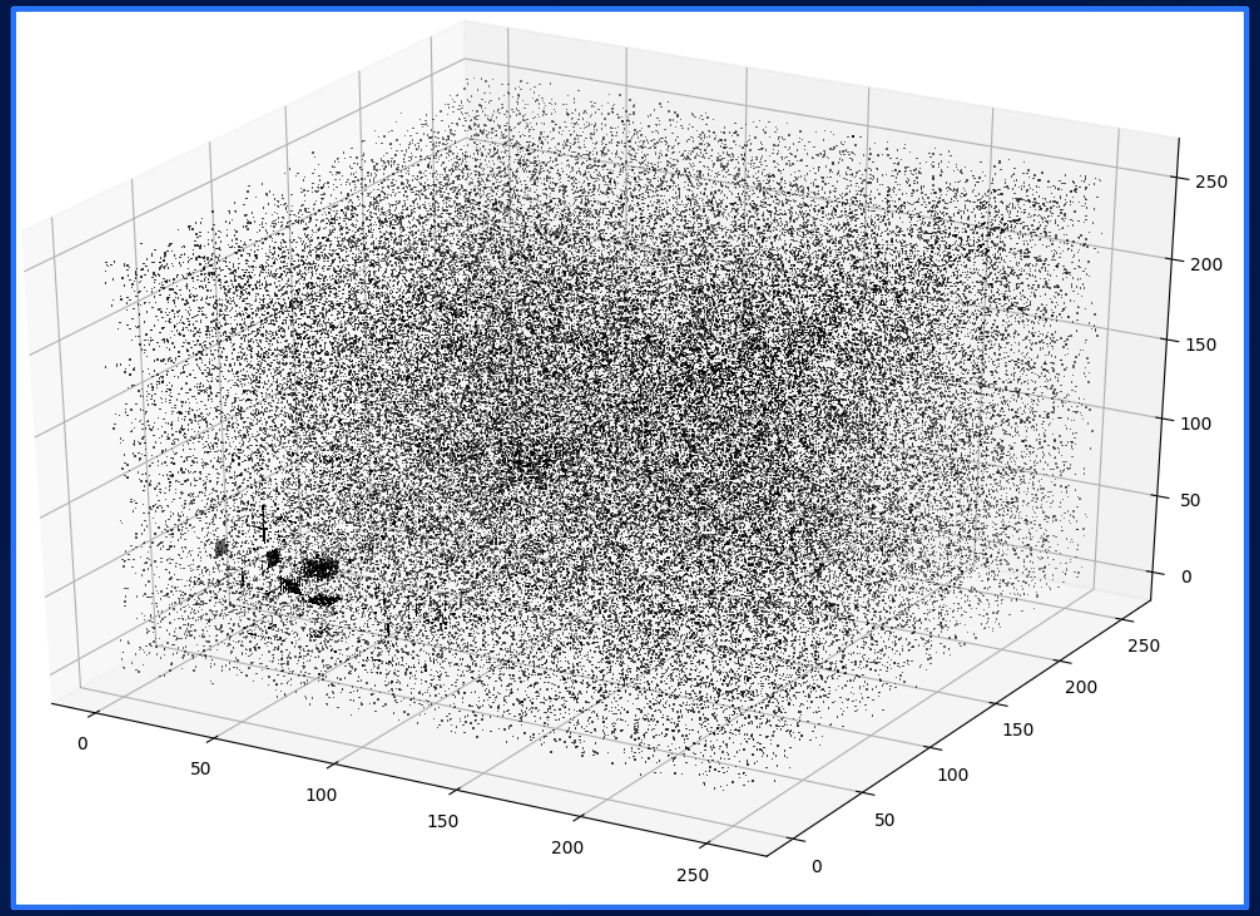
DATA CONTEXT

calculator

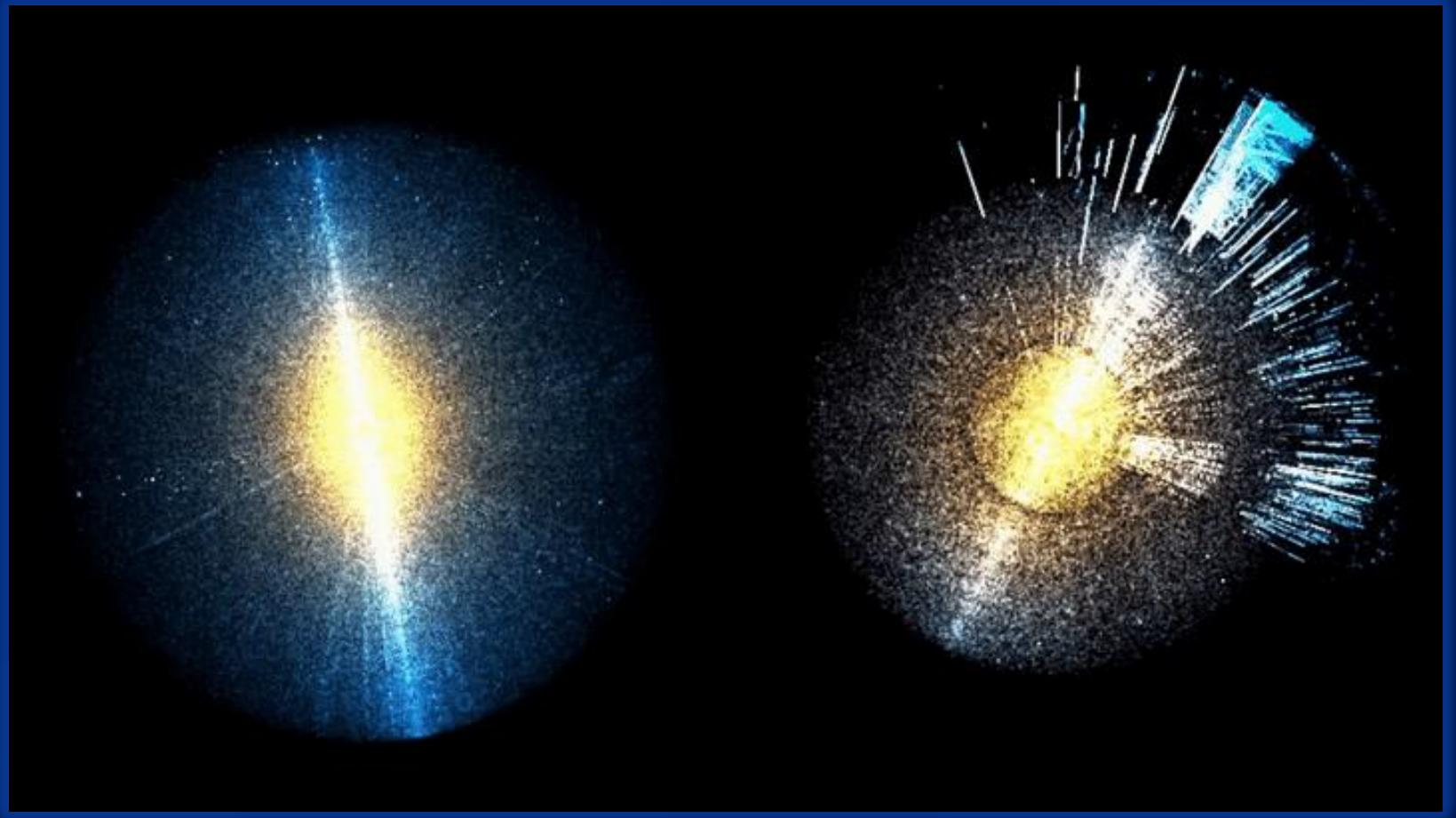
Steuererklärung (PDF)



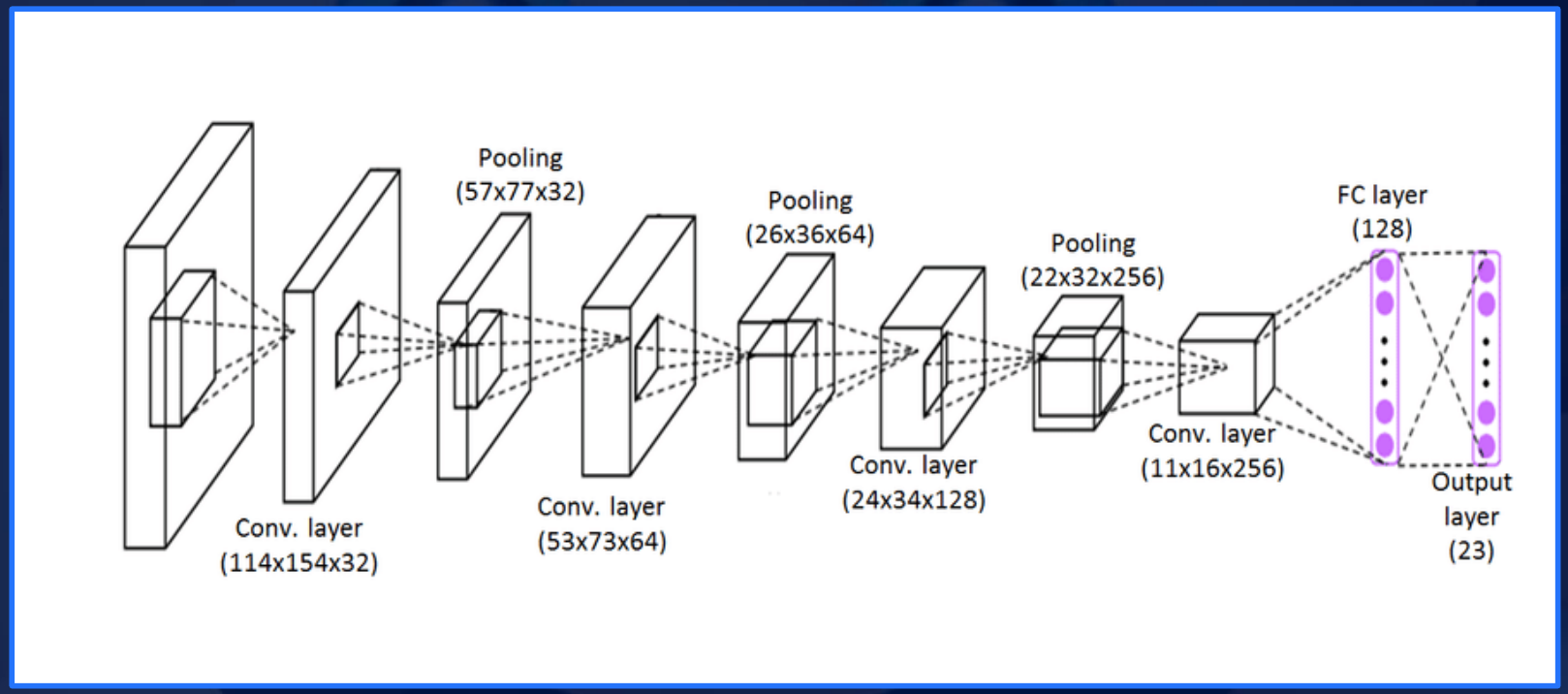
DATA CONTEXT



DATA
CONTEXT



CNN
CONCEPT



CONFLICT! -> Data reduction required

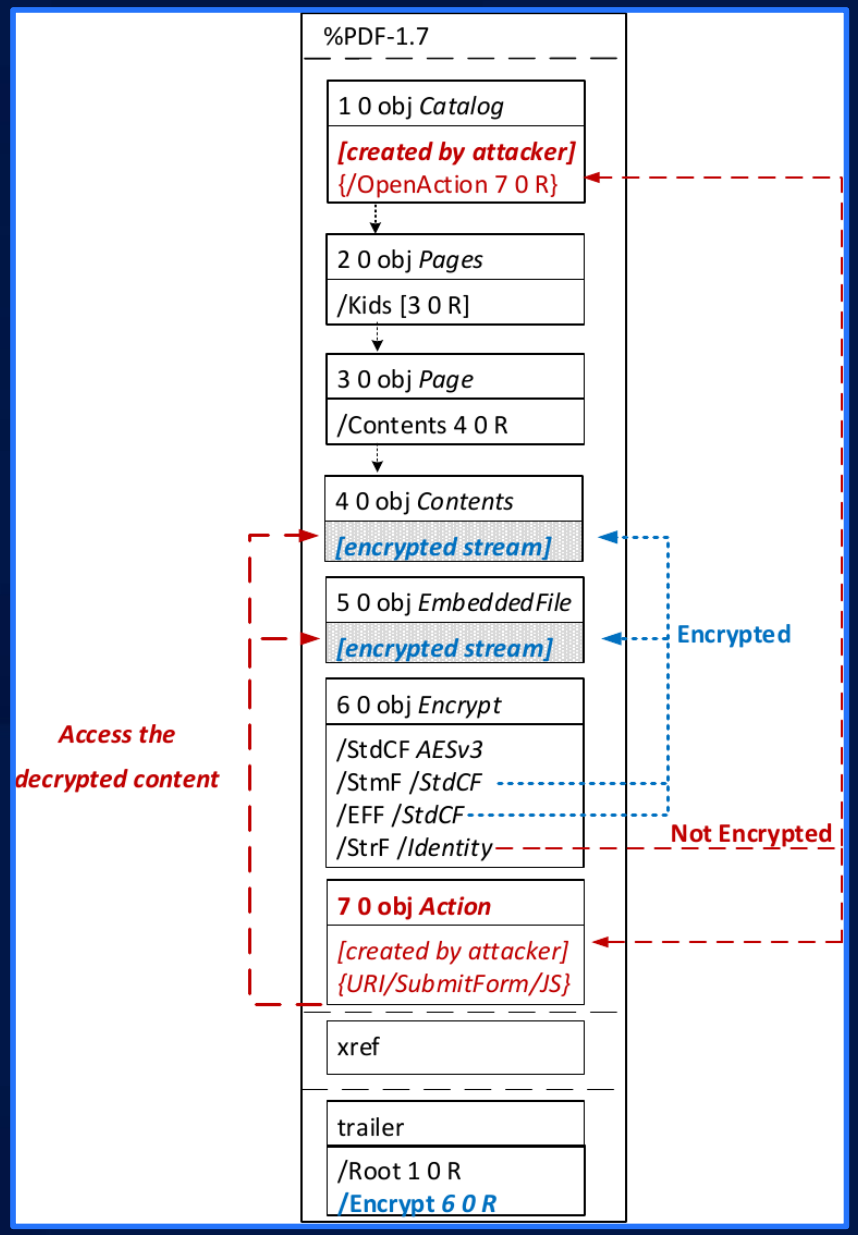
DATA SELECTION

Multitude of different information

- Context, Locality, Randomness, etc.

Selection, Segmentation & Reduction

- Structural Information
- User Data
  - Attachments
  - Encryption
- Encodings
- Filters



INLYSE

DATA OBJECTIVES

Multitude of different information

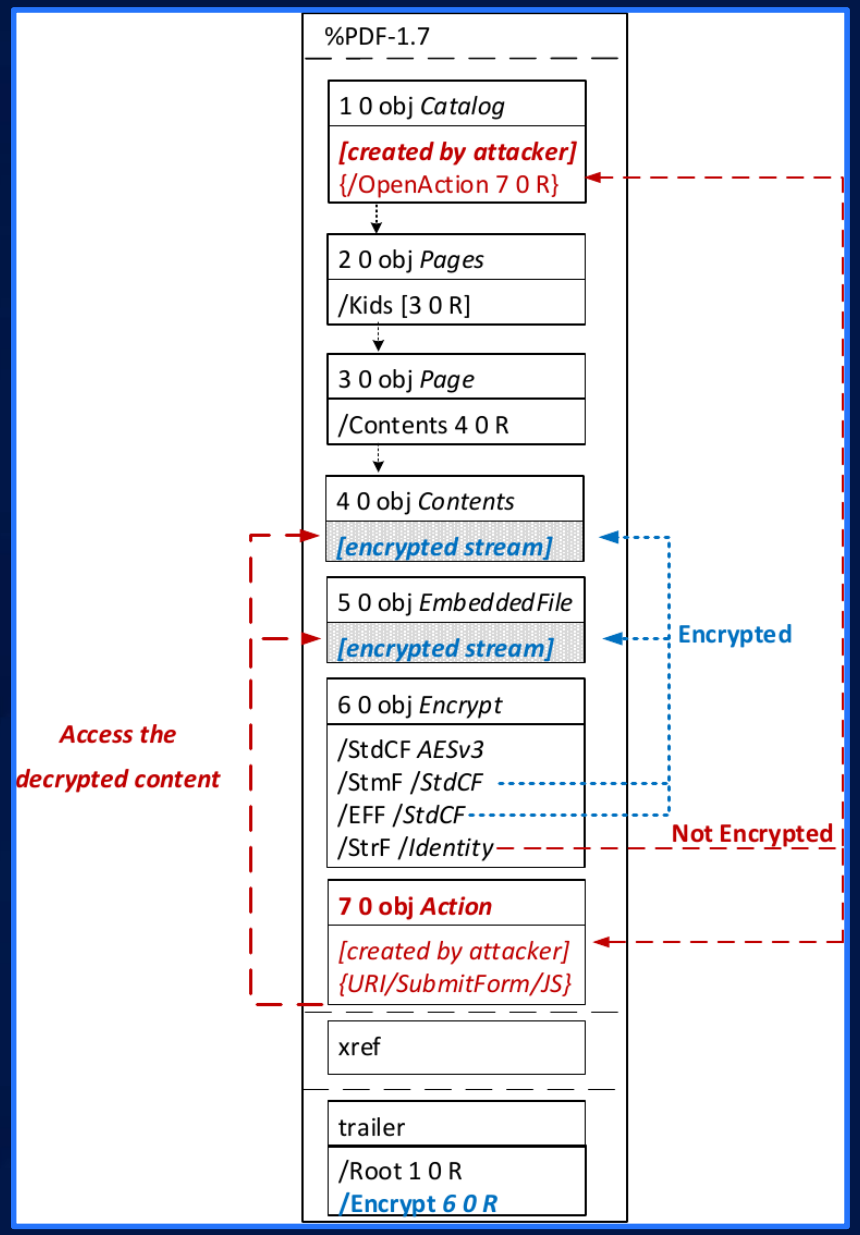
Context, Locality, Randomness, etc.

Selection, Segmentation & Reduction

Decisions on how to deal with which information

Representation

Different conversions for different information



INLYSE

DETECTION RATES
TAKE AWAYS

DATASET

	PDF	MS-Office
benign	12.399.298	9.637.553
malicious	3.609.048	3.209.806
Total	**16.008.346**	**12.847.359**

Sources
Virusshare
Virustotal
Malshare
Self collected
Cooperations

DETECTION RATES

	PDF	MS-Office
True Positive Rate	99,59492%	99,54643%
False Positive Rate	0,08819%	0,09383%
True Negative Rate	99,91181%	99,90616%
False Negative Rate	0,05623%	0,05761%

AVERAGE ANALYSIS TIME

3s

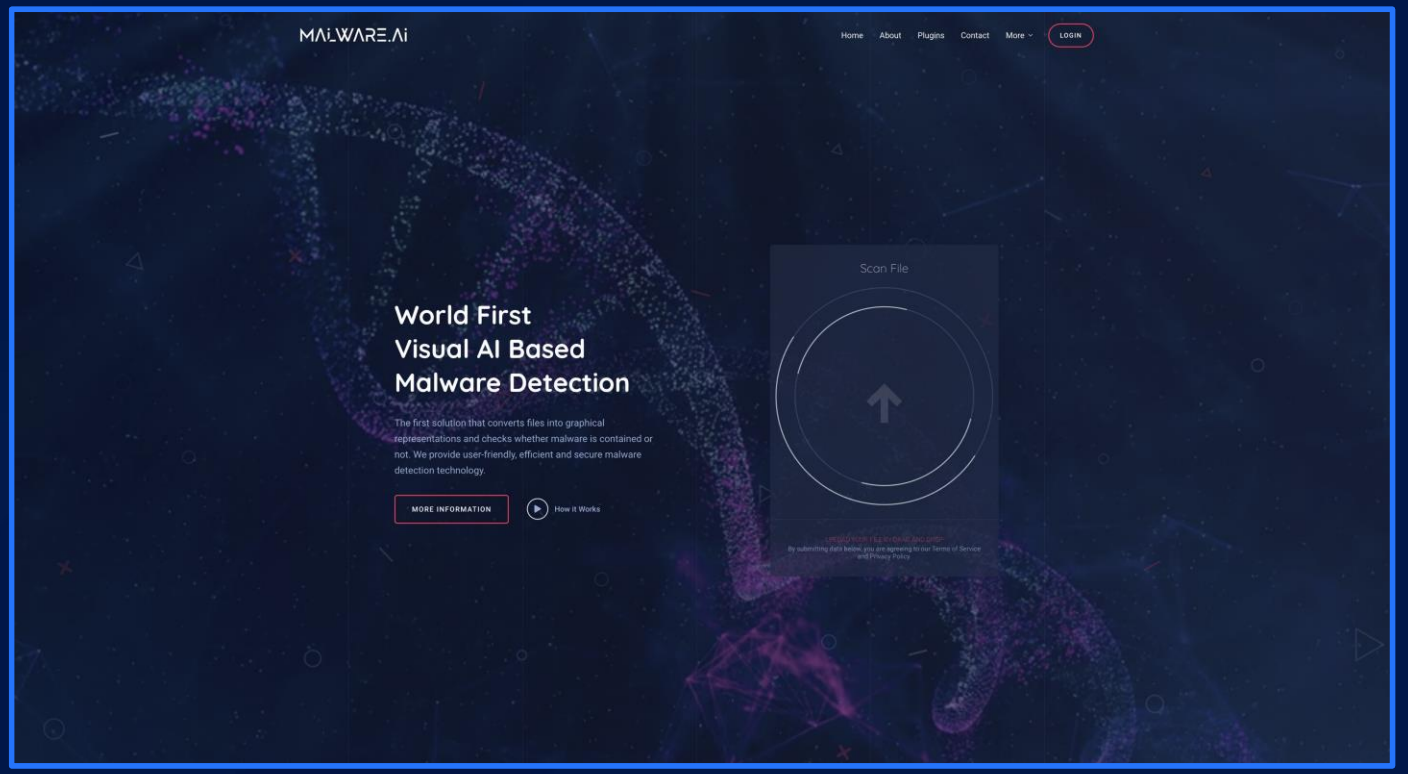
UPDATE CYCLE

>8 month

INLYSE

THANKS FOR

WATCHING



INLYSE

CONTACT US

www.inlyse.com

info@inlyse.com