

MLPdf:

An Effective Machine Learning Based Approach for PDF Malware Detection

Dr Jason Zhang

Senior Threat Researcher

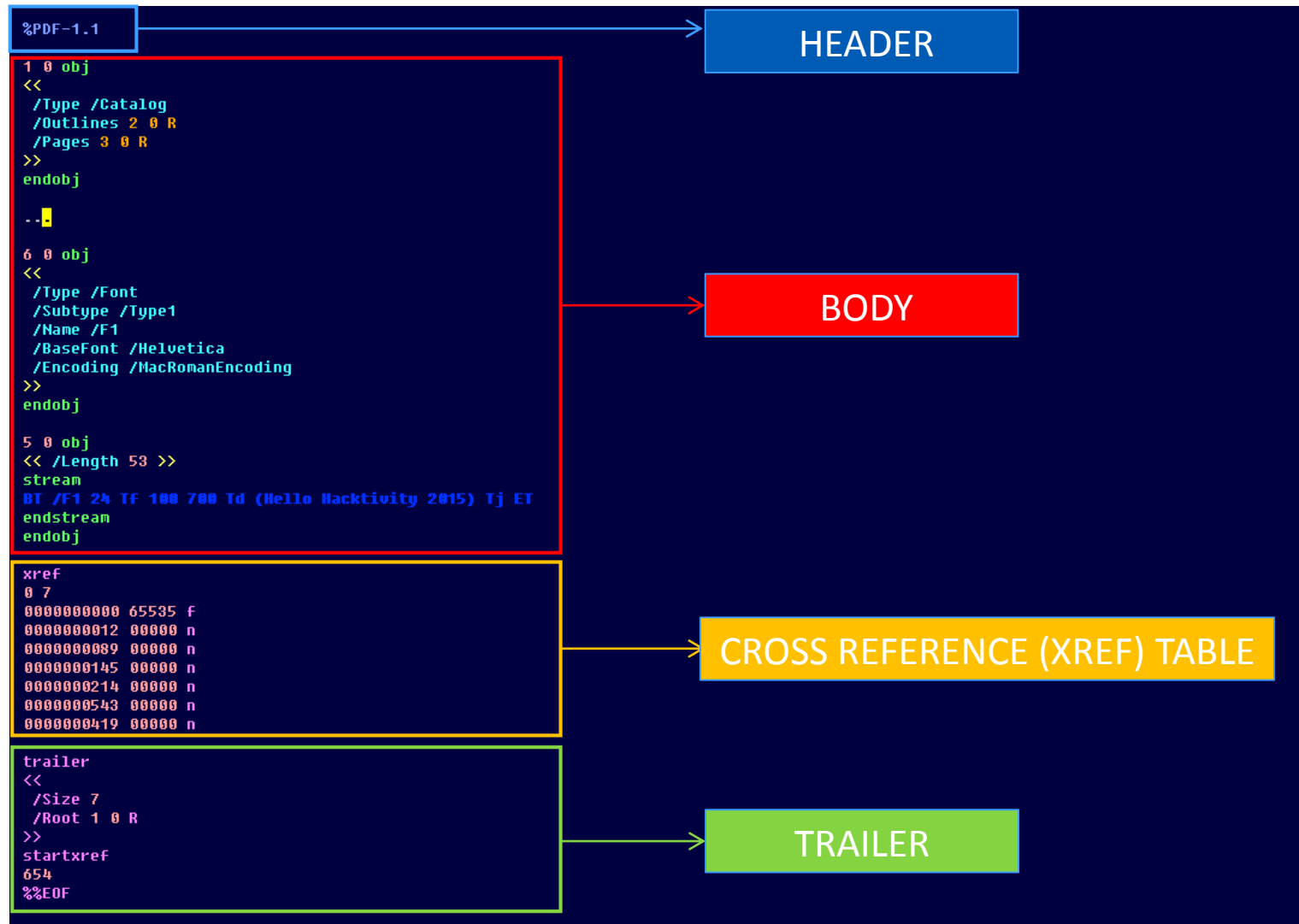
SOPHOS

Agenda

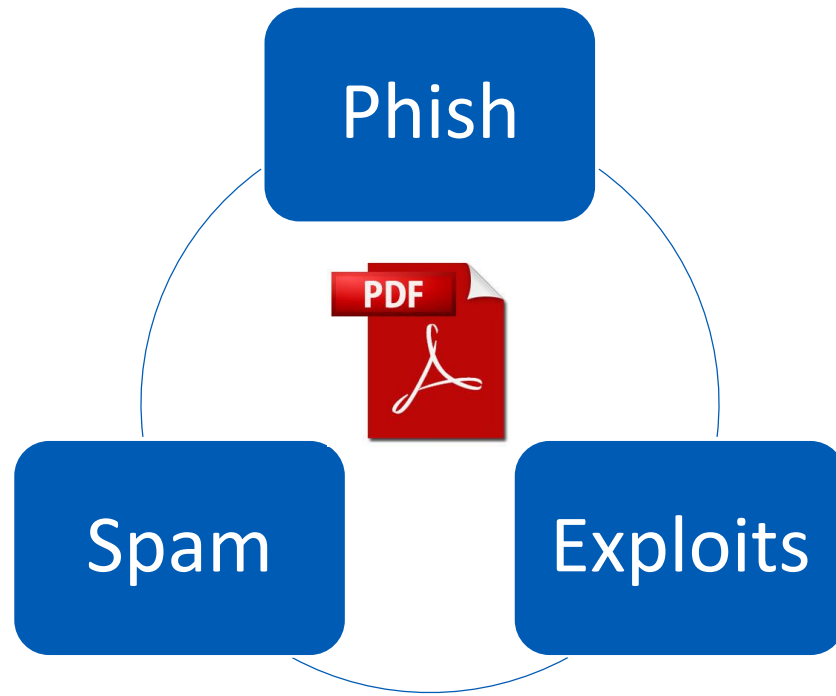
- PDF File Format and Associated Attacks
- MLPdf Model
- Data Labelling
- Feature Engineering
- Underfitting & Overfitting
- Demo
- Summary
- QA

PDF File Format & Associated Attacks

PDF File Format

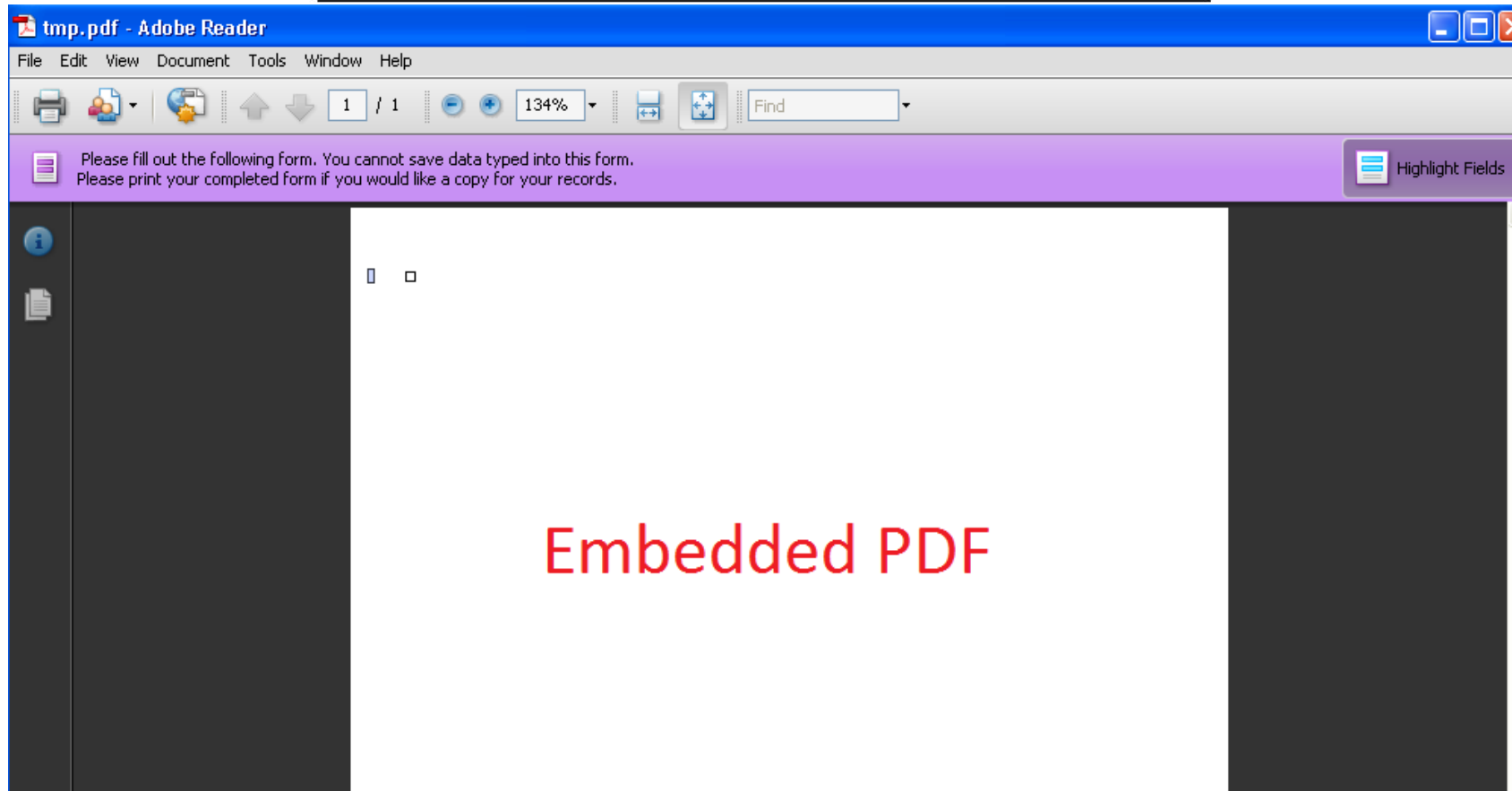


PDF Based Attacks



PDF Based Attacks: Exploits

Film Chip Capacitors
PEN DIELECTRIC – CB Series



%PROFILE%\Local Settings\Temp\A9RDC52.tmp\tmp.pdf

PDF Based Attacks: Exploits (cont.)

a)

```
obj 241 0
Type:
Referencing: 242 0 R
<<
  /Names [ <tmp.pdf> 242 0 R ]
>>

obj 242 0
Type: /Filespec
Referencing: 243 0 R
<<
  /F 243 0 R
>>
```

XFA oneOfChild (CVE-2013-0640)

embedded.pdf

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0123456789ABCDEF
0000h:	25	50	44	46	2D	31	2E	32	0D	0A	31	20	30	20	6F	62	%PDF-1.2...1 0 ob
0010h:	6A	0D	0A	3C	3C	0D	0A	09	2F	50	61	67	65	73	20	32	j...<<.../Pages 2
0020h:	20	30	20	52	0D	0A	09	2F	41	63	72	6F	46	6F	72	6D	0 R.../AcroForm
0030h:	20	3C	3C	0D	0A	09	09	2F	58	46	41	20	34	20	30	20	<<.../XFA 4 0
0040h:	52	0D	0A	09	3E	3E	0D	0A	09	2F	5A	5A	5A	20	3C	3C	R...>>.../ZZZ <<
0050h:	0D	0A	09	09	2F	45	45	45	20	35	20	30	20	52	0D	0A	.../EEE 5 0 R..
																	.>>.../OpenActio
																	n <<.../JS (str
																	= "ZnVuY3Rpb24g
																	c0hPR0coYyxkLGUp
																	ewogICAgdmFyIGlk
																	eCA9IGQgJSBjImxl
																	bmd0aDsKICAgIHZh
																	ciBzID0gIiI7CiAg
																	ICB3aGlsZSAocy5s

b)

```
/ViewerPreferences
<<
  /Direction /L2R
>>

obj
Ty
Re
Co
<
>
```

Util.printf() (CVE-2008-2992)
Collab.getIcon (CVE-2009-0927)
Collab.collectEmailInfo (CVE-2007-5659)
Escript.api plugin media player (CVE-2010-4091)

embedded_obj191.pdf

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	0123456789ABCDEF
0300h:	72	65	61	6D	0D	0A	65	6E	64	73	74	72	65	61	6D	0D	ream..endstream.
0310h:	0A	65	6E	64	73	74	72	65	61	6D	0D	0A	65	6E	64	73	.endobj..76 0 ob
																	j...<</S/JavaScri
																	pt/JS(j='vt34t';
																	..a=new String(\
																	"if\ (e'1\)\bjsg=
																	u836c45790d2a;nt
																	o zvr,qy{whl.*<+
																	}/xkAp_-'[]SCEMI
																	:WNKQDUMGPV &>@"
																	\);..b="1";..b2=
																	"a"..+b;..z='0*1
																	*2*3*2*4*5*4*6*6
																	*7*8*9*10*11*4*1
																	2*13*14*15*16*16

PDF Based Attacks: Phish



PDF Based Attacks: Spam



Объявляем распродажу

Е-мейл рассылок!

Едино разовая рассылка
по 1 любому городу за

~~7000~~ 3500 рублей!

Всего 3500 рублей и
Ваше рекламное письмо
получат 1 088 958* адресат

Punycode
based spam

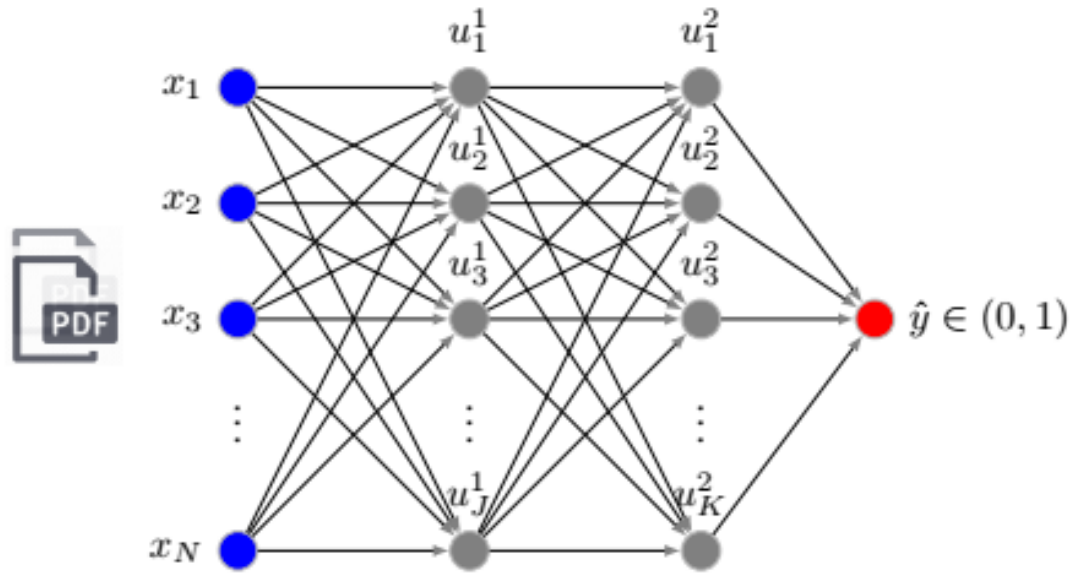


**кол-во адресатов указано для города Москва,
ознакомится с полным списком городов
и количеством адресатов в каждом*

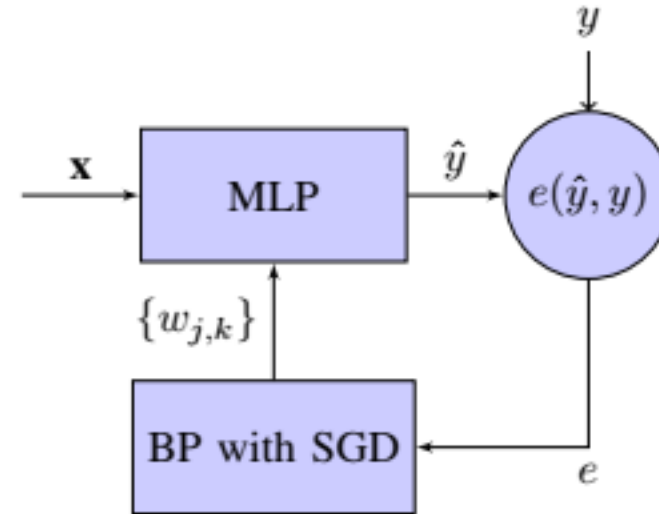
[МОЖНО НА САЙТЕ](#)

MLPdf Model

MLPdf Model



MLP neural network model



BP based MLP weights update via SGD

$$\Delta w_{j_{i-1}, k_i} = -\eta \frac{\partial e(\hat{y}, y)}{\partial w_{j_{i-1}, k_i}}$$

8857 parameters (weights + bias terms) to be updated

Data Labelling

Data Labelling

Know your **enemy**



It is critical to accurately label **clean** and **malicious** files to train ML algorithms.

Data Labelling

A file is labelled as **MALWARE** if detected by

- ≥ 3 reliable AV scanners, OR
- ≥ 1 trustable SAV identities

A file is labelled as **CLEAN** if

- not detected by any scanner, OR
- confirmed as FP

Feature Engineering

Feature Engineering



PART OF THE EXTRACTED FEATURES

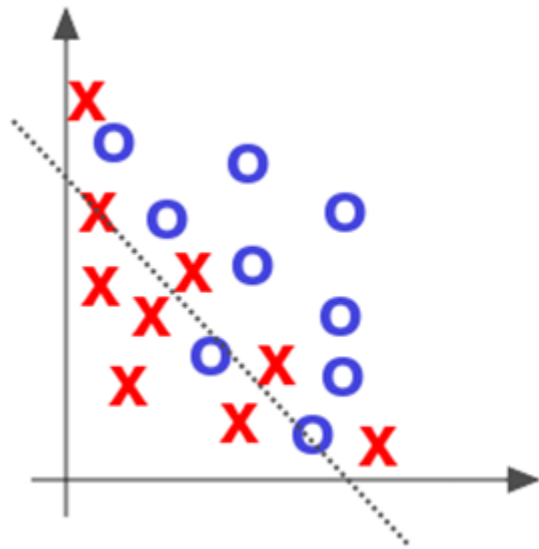
Feature name	Description
<i>F_SIZE</i>	<i>PDF file size</i>
<i>F_JS</i>	<i>PDF with JavaScript or not</i>
<i>F_PGC</i>	<i>Page count</i>
<i>F_OBJC</i>	<i>Number of objects</i>
<i>F_FILT</i>	<i>Stream filtering</i>
<i>F_ENTRP1</i>	<i>Entropy of some content</i>
<i>F_ENTRP2</i>	<i>Entropy of some content</i>
...	...

In-house & 3rd party tools to extract **48** features

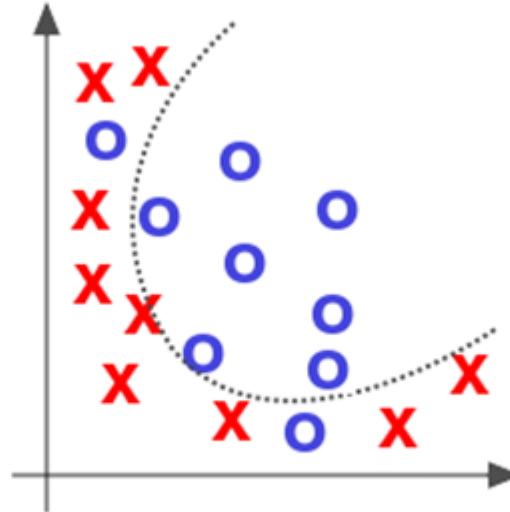
Underfitting & Overfitting

Underfitting & Overfitting

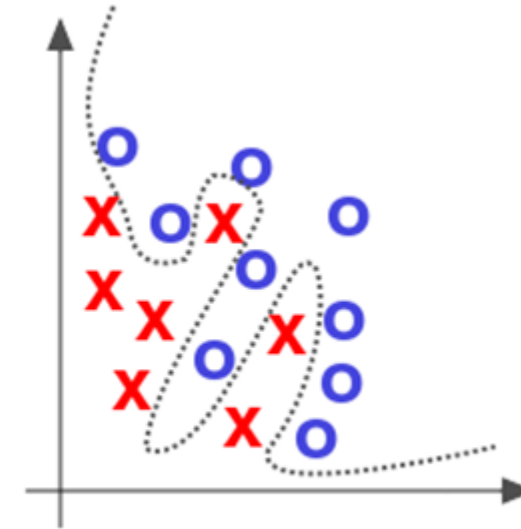
- Make the training error small
- Make the gap between training and test error small



Underfitting



Good fit/Robust



Overfitting

a) Validation b) Batch normalization c) Dropout

Demo

Demo

- *Demo-1*: clean (13101) & malware (1946) files


Source: Sophos filesDB & URL logs prior to March 31, 2018

- *Demo-2*: clean files (35599)

Source: Sophos URL logs for May 9-24, 2018

- *Demo-3*: a case study with a zero-day attack

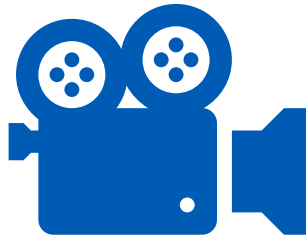
Source: Sophos filesDB - July 2018



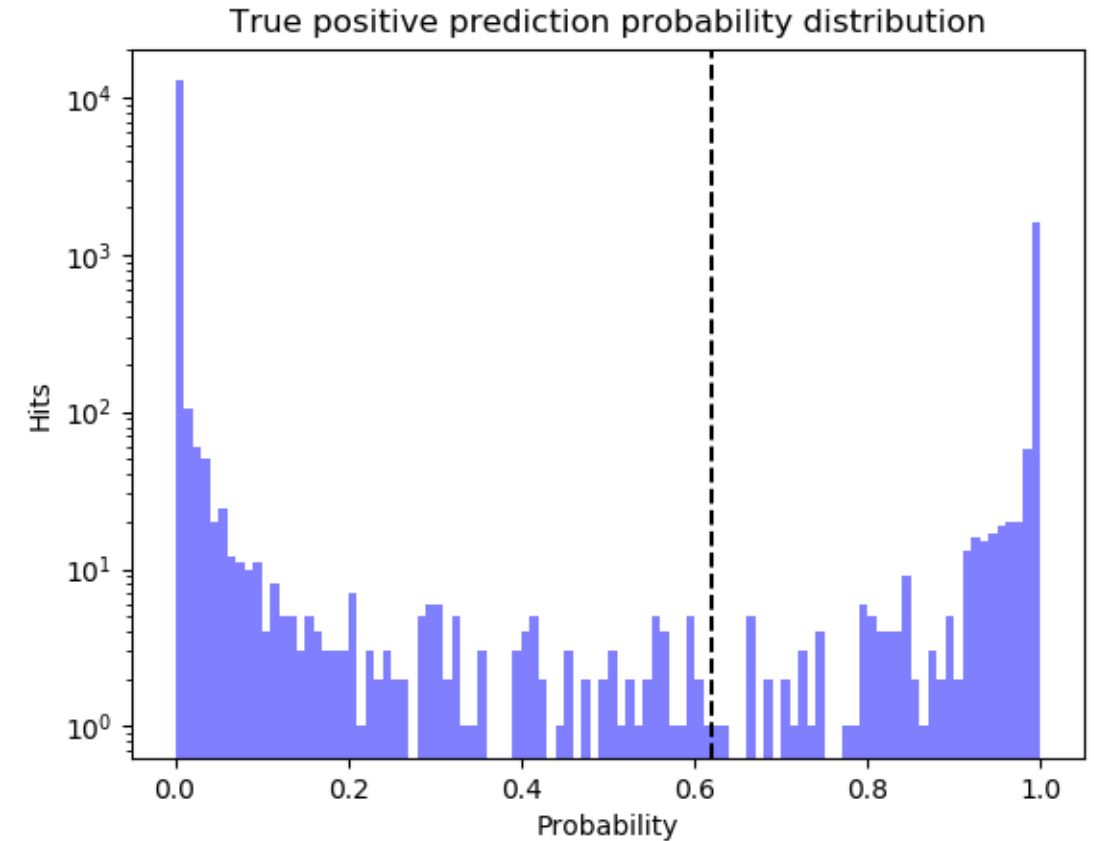
File size
< 15MB

Demo: Dataset-1 – Clean (13101) & Malware (1946) Files

Source: Sophos filesDB & wild collections prior to March 31, 2018

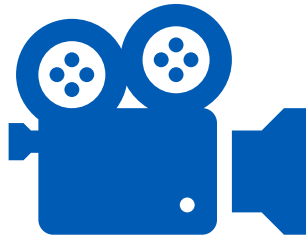


	FPR	TPR
MLPdf	0.12%	95.38%

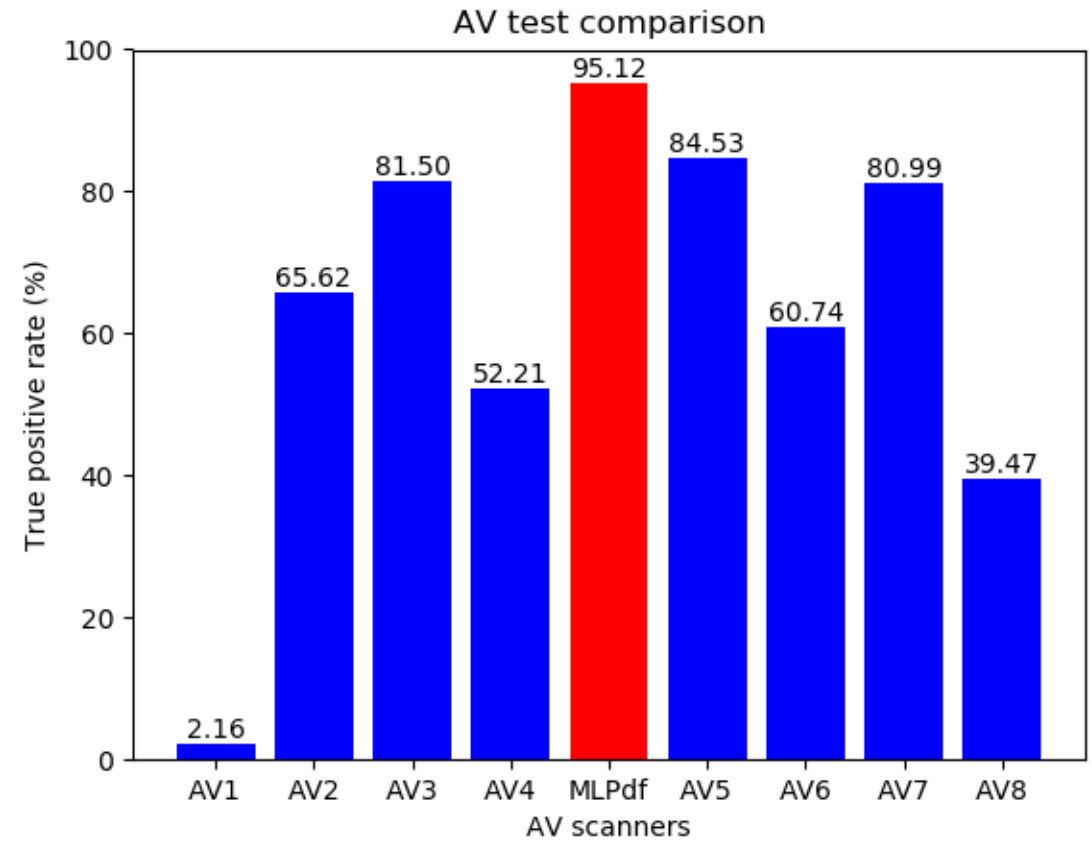


Demo: Dataset-1 (cont.) – Compared with 8 AV Scanners

Source: Sophos filesDB & wild collections prior to March 31, 2018

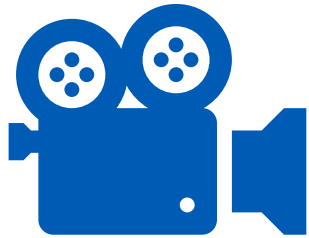


	FPR	TPR
MLPdf	0.12% (0.08%)	95.38% (95.12%)

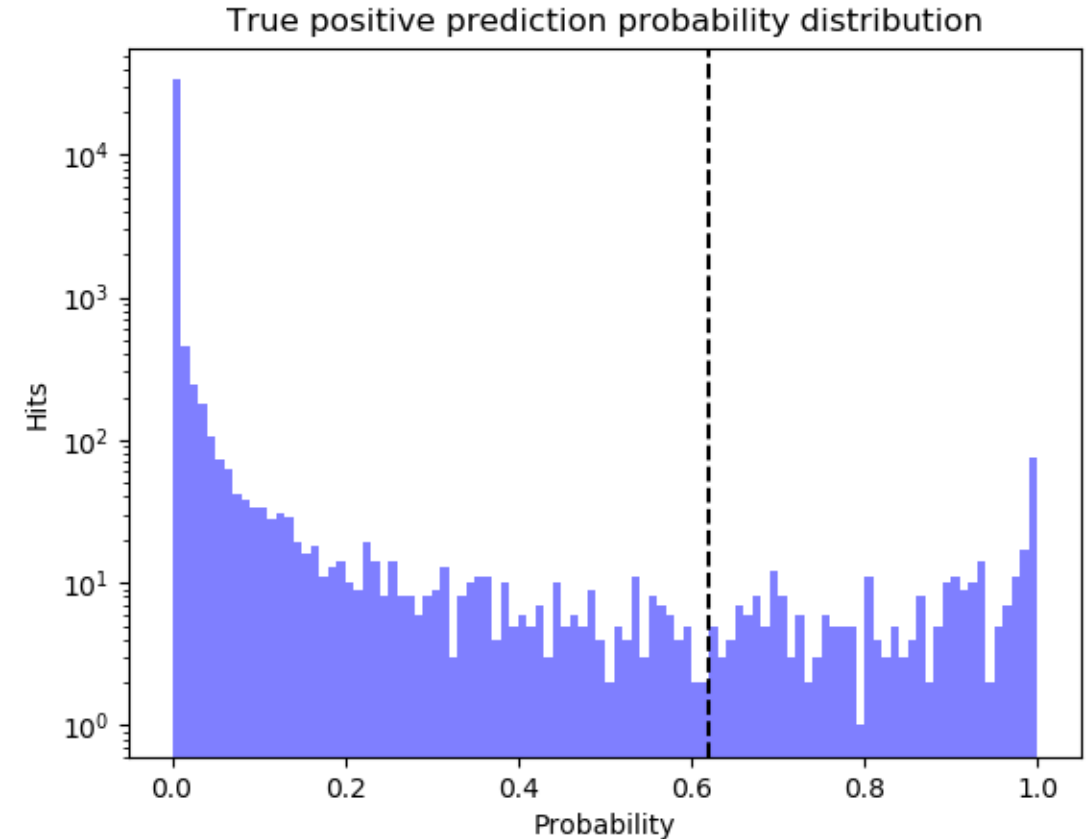


Demo: Dataset-2 – Clean Files (35599)

Source: SXL3 URL logs for May 9-24, 2018

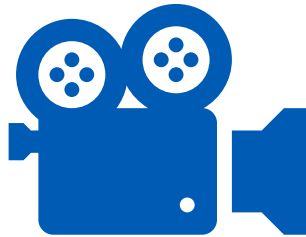


	FP	FPR
MLPdf	311	0.87%

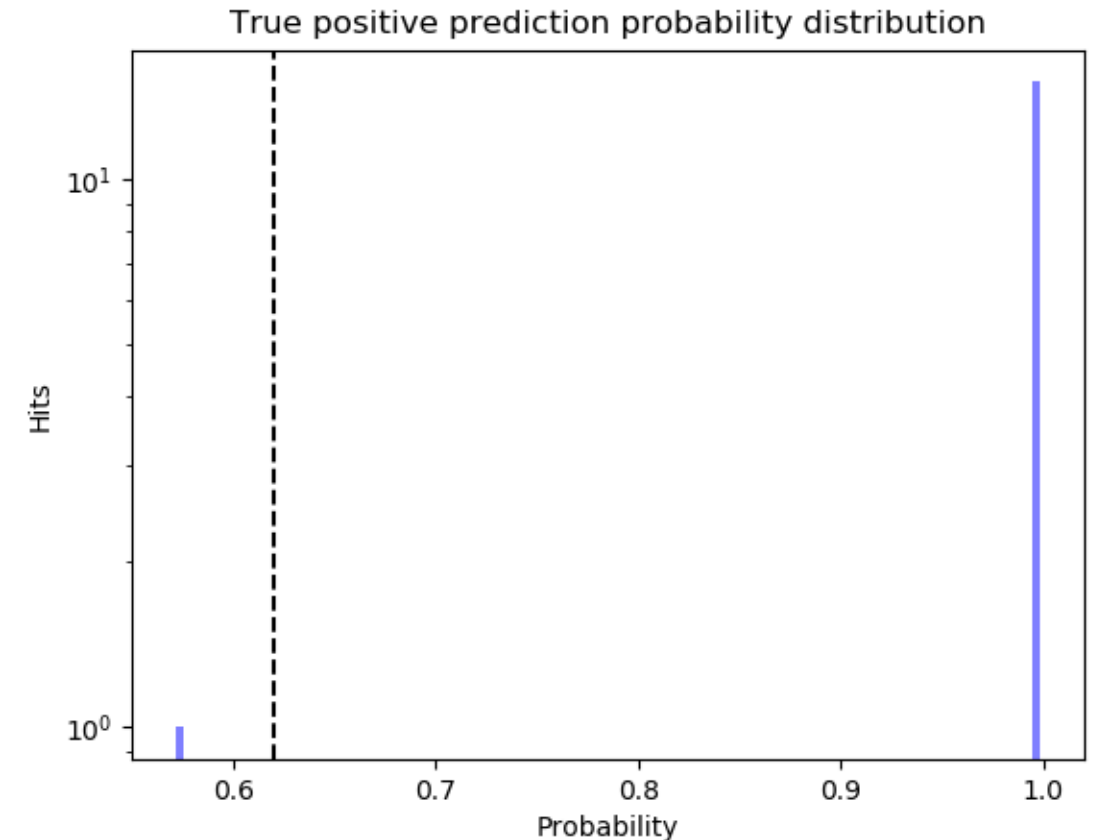


Demo: Dataset-3 – A Case Study with a Zero-Day Attack

Source: Sophos filesDB – July, 2018



	FN	TPR
MLPdf	1	93.75%



Summary

- Overview of PDF file format and typical attacks
- Introduction to MLPdf model
- Feature engineering & general challenges in ML
- MLPdf demo with various datasets, including analysis for a zero-day attack
- Pros & Cons of MLPdf vs traditional AV scanners

QA



Thank You

White paper: <https://arxiv.org/abs/1808.06991>

Email: jason.zhang@sophos.com

SOPHOS
Security made simple.