black hat USA 2016

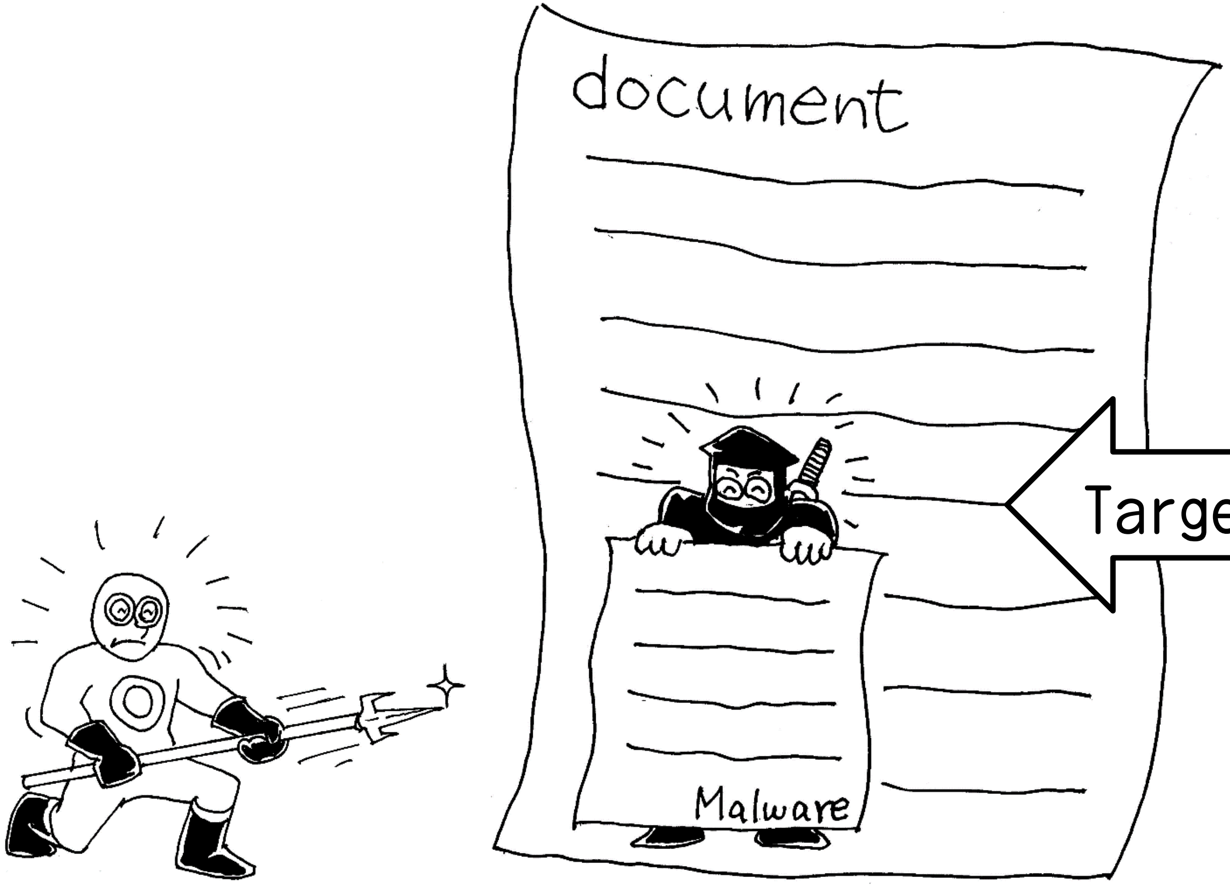
JULY 30 – AUGUST 4, 2016 / MANDALAY BAY / LAS VEGAS

O-checker:
Detection of Malicious Documents through Deviation from File Format Specifications

Yuhei Otsubo

Agenda

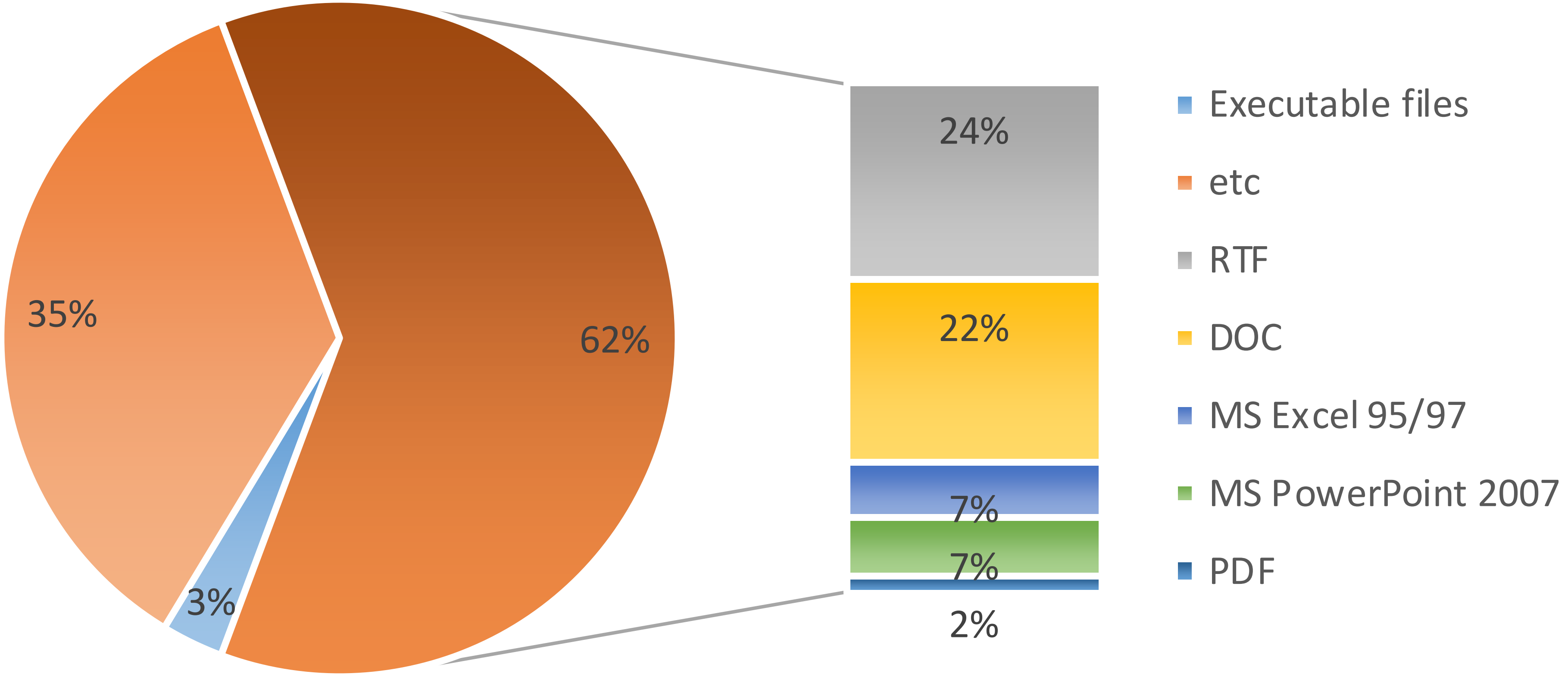
1. Overview of o-checker
2. (DEMO)How to use o-checker



Targeted email attacks

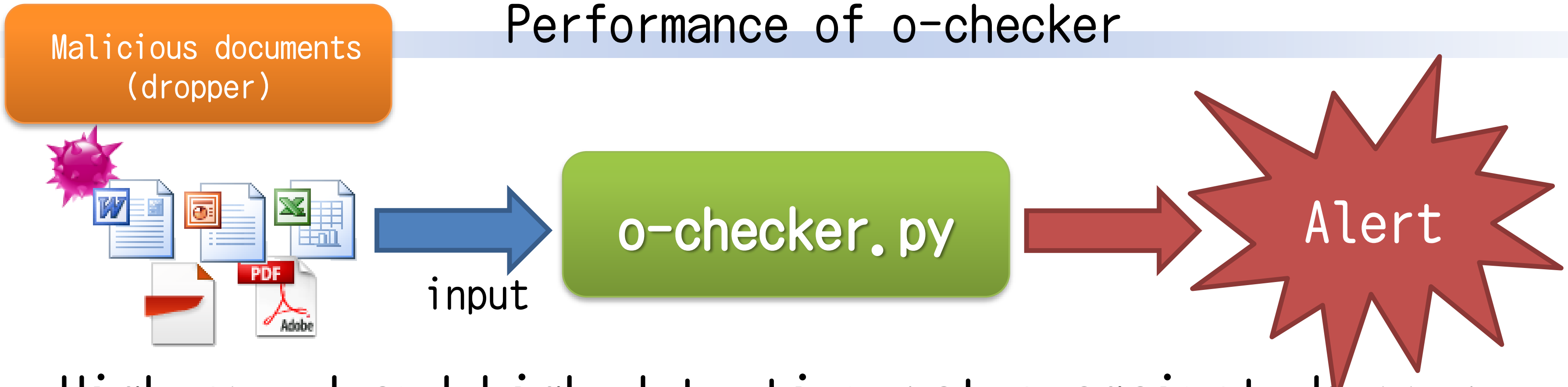
Attachment files in targeted email attacks in 2014

Over 60% of the attachment files are document files



according to “TrendLabs 2014 Targeted Attack Campaign Report”

Performance of o-checker



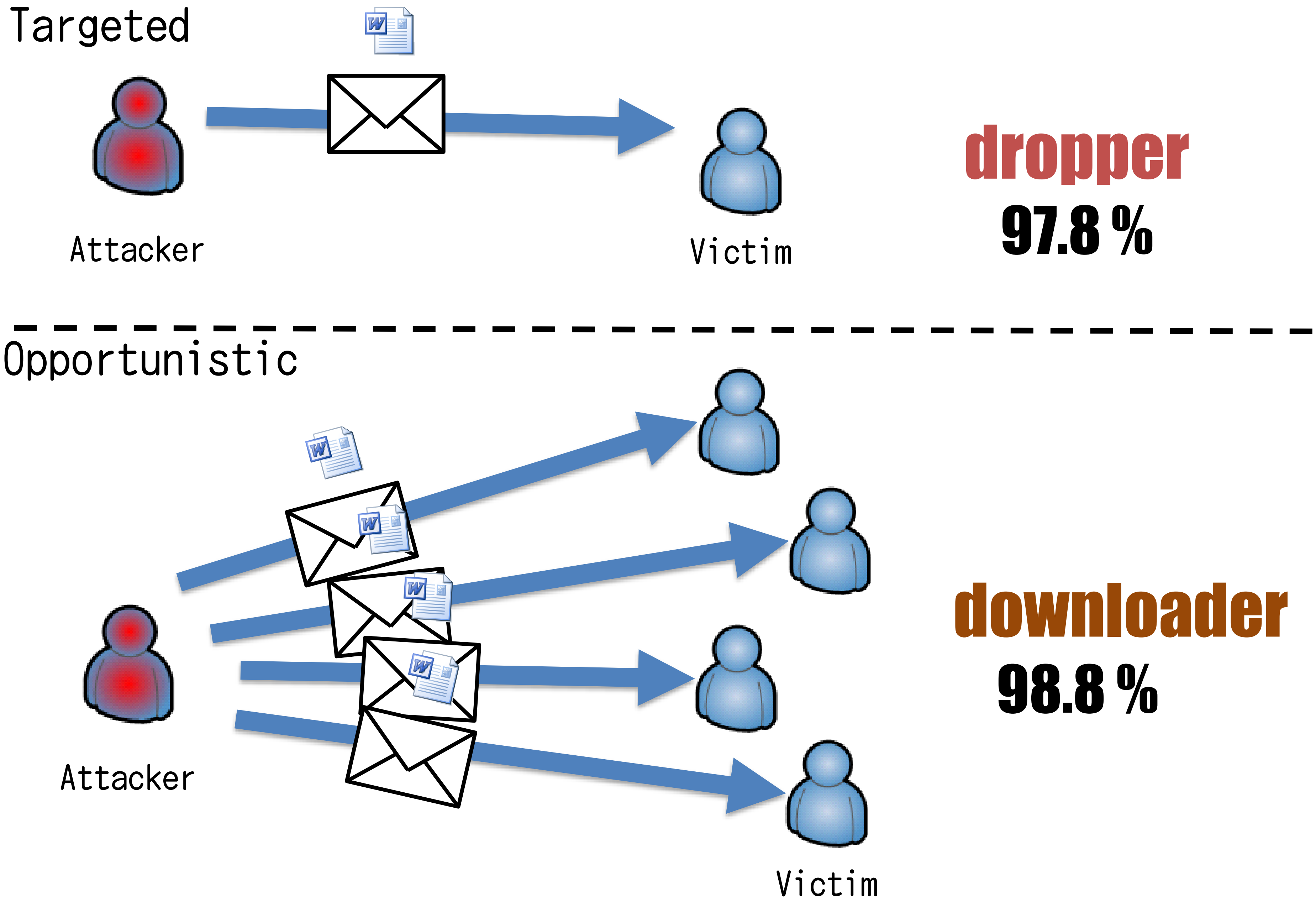
- High speed and high detection rates against dropper
  TPR 2009-2012:**99.2%**(360/363) FPR **0.3%**(35/10,801)
  2013-2014:**98.4%**(122/124)
  Average execution time:**0.3 sec**
- Almost maintenance-free

We have **never changed** the detection methods **since Apl.2013**.

	Updating frequency	Remarks
Anti-virus software	Every day	310,000 new type of malware per day (2015)※
o-checker	**Almost none**	It needs update, if a new document file format comes out.

※：http://usa.kaspersky.com/about-us/press-center/press-releases/new-daily-malware-count-kaspersky-lab-decreases-15000-2015

Trend of malicious documents



Targeted

dropper

97.8 %

Opportunistic

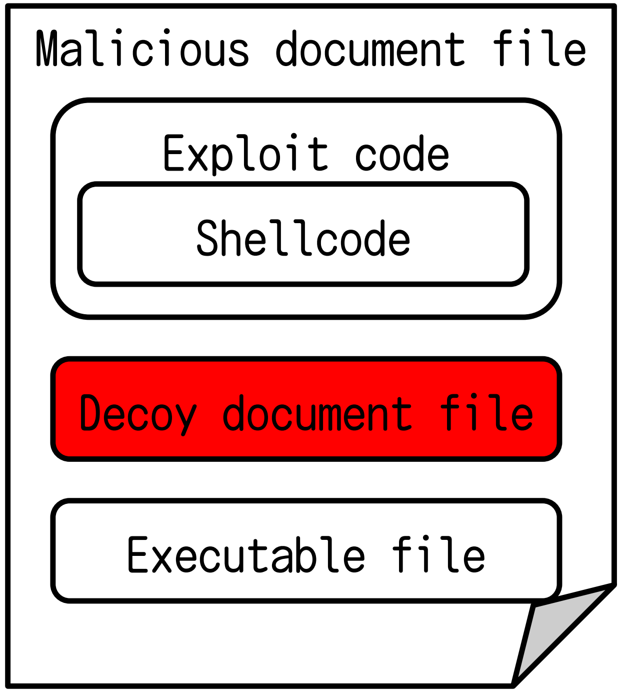
downloader

98.8 %

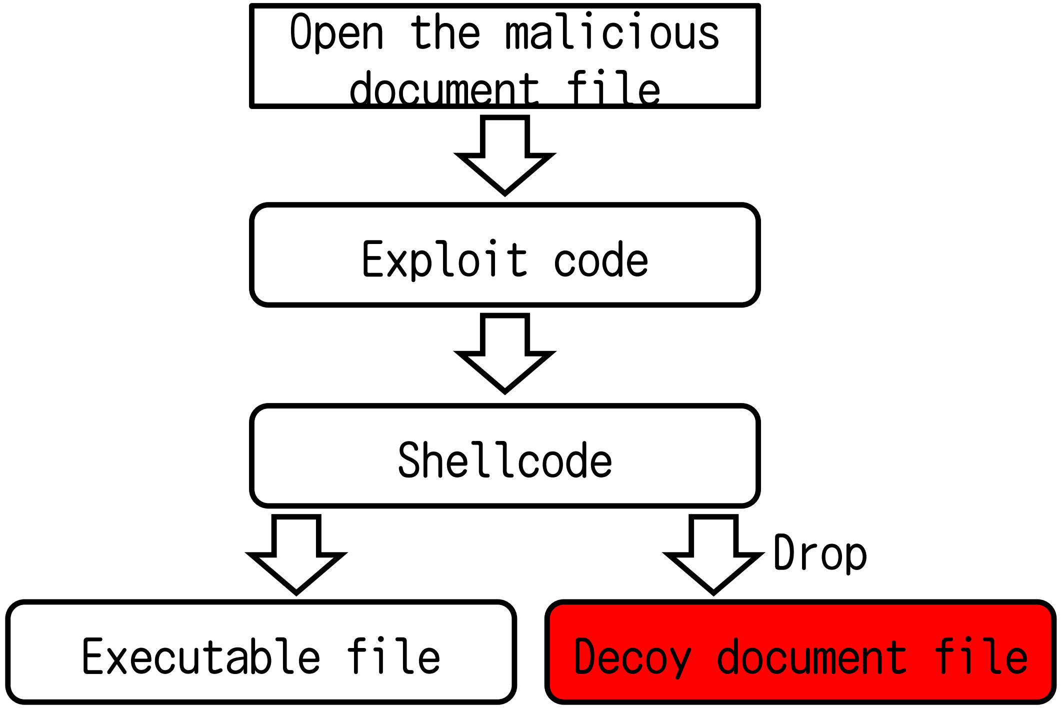
Why dropper?

Victims **consciously open** malicious documents

Typical structure



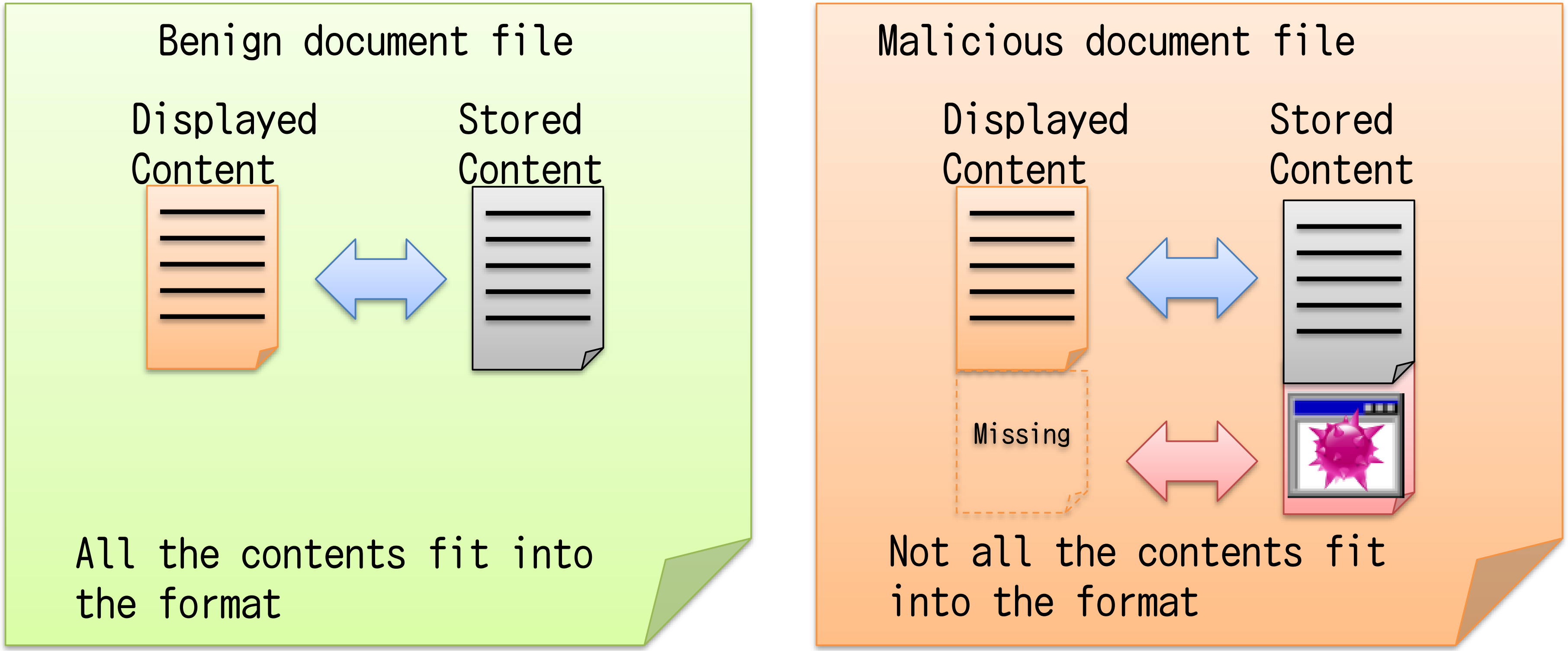
Typical execution process



Background

Foreground

Detection mechanism (simplified)



Benign document file

Displayed Content

Stored Content

All the contents fit into the format

Malicious document file

Displayed Content

Stored Content

Not all the contents fit into the format

“o-checker” checks the anomaly structure of a malicious document file

Overview of tar(09-12)

We examined various document files used in targeted attacks from 2009 to 2012.

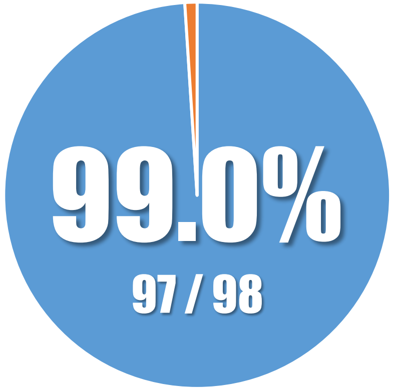
File type	Ext.	Num.		Avg. size(KB)
		dropper	downloader	
RTF	rtf	**98**	1	266.5
CFB	doc	**36**	0	252.2
	xls	**49**	0	180.4
	jtd/jtdc	**17**	0	268.5
PDF	pdf	**163**	7	351.2
Total	Num.	**363**	8	291.8
	Rate	**97.8 %**	2.2 %	

- tar(09-12) were used in targeted email attacks from 2009 to 2012
- Most of all the files are droppers

※ “jtd/jtdc” file type is used in Japanese Word Processor named “一太郎” (Ichitaro).

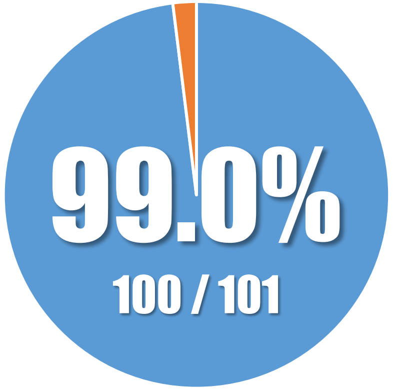
Rate of each anomaly structure

We classified **8 anomaly structures**.
We can classify **99.2%**(360/363) of the droppers of tar(09-12) according to these features.



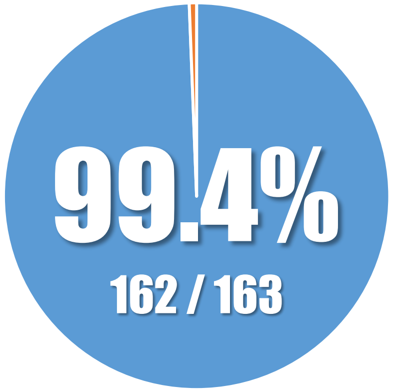
RTF

AS1 99.0%



CFB

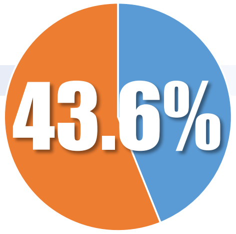
AS2 78.2%
AS3 91.1%
AS4 98.0%
AS5 97.0%

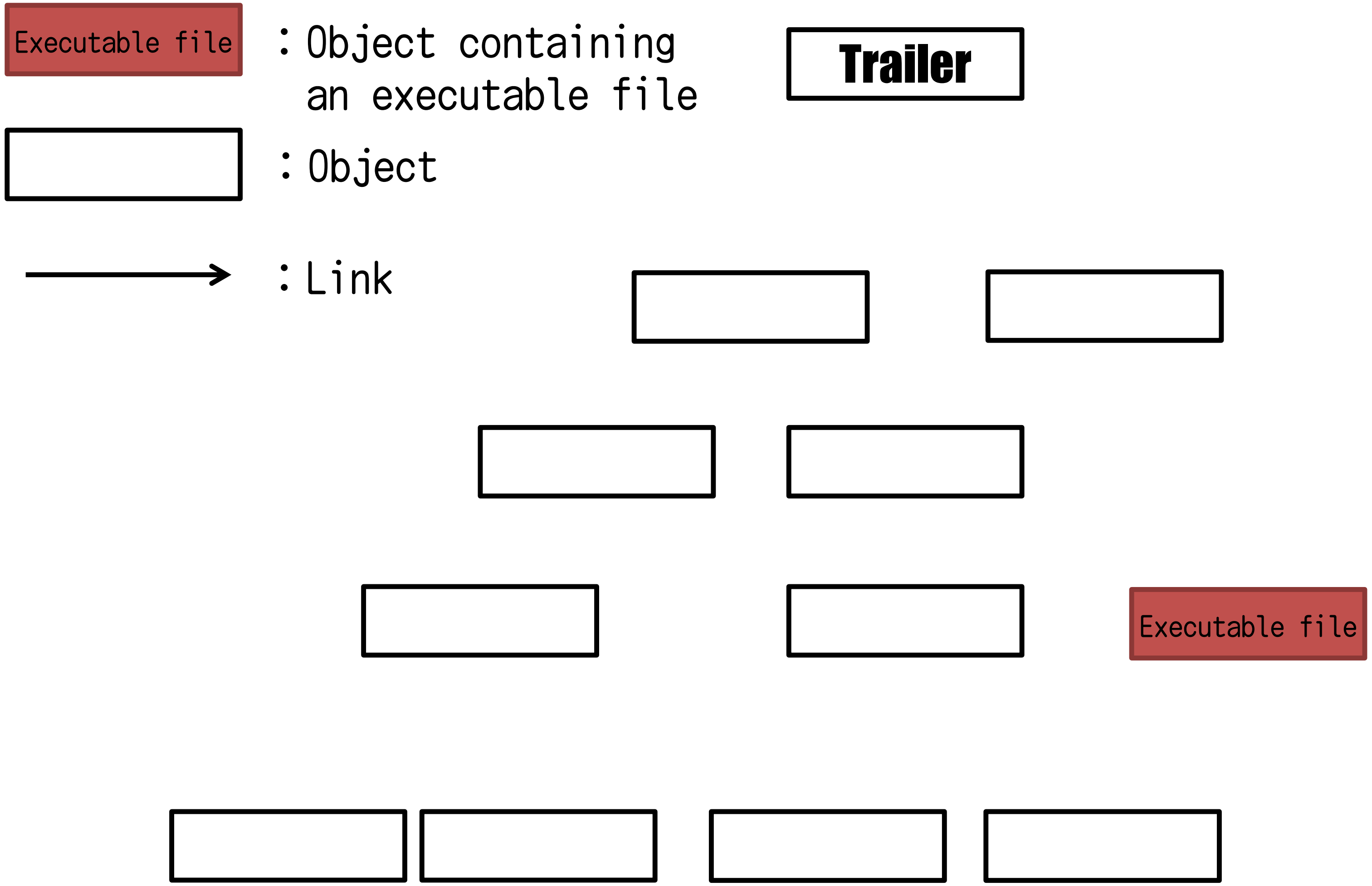


PDF

AS6 49.7%
AS7 43.6%
AS8 62.6%

AS7:Unreferenced object

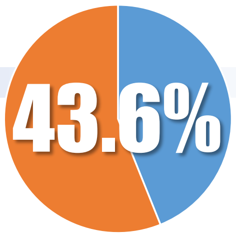


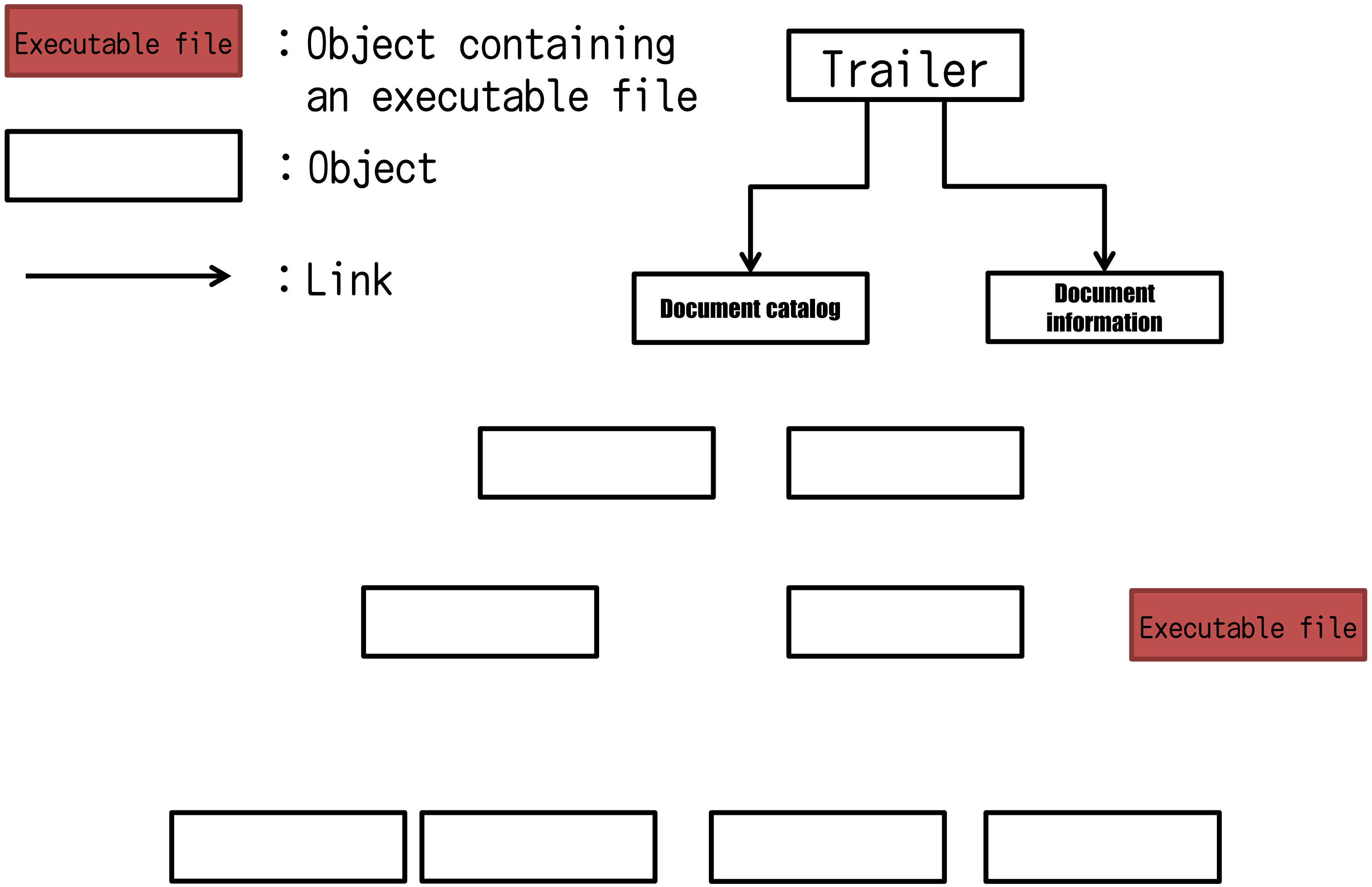


A PDF file containing an executable file

When an executable file is inserted as an object in disregard of document structure, it is often unreferenced.

AS7:Unreferenced object

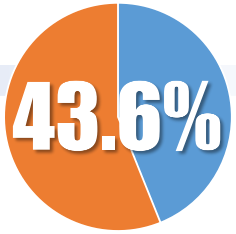


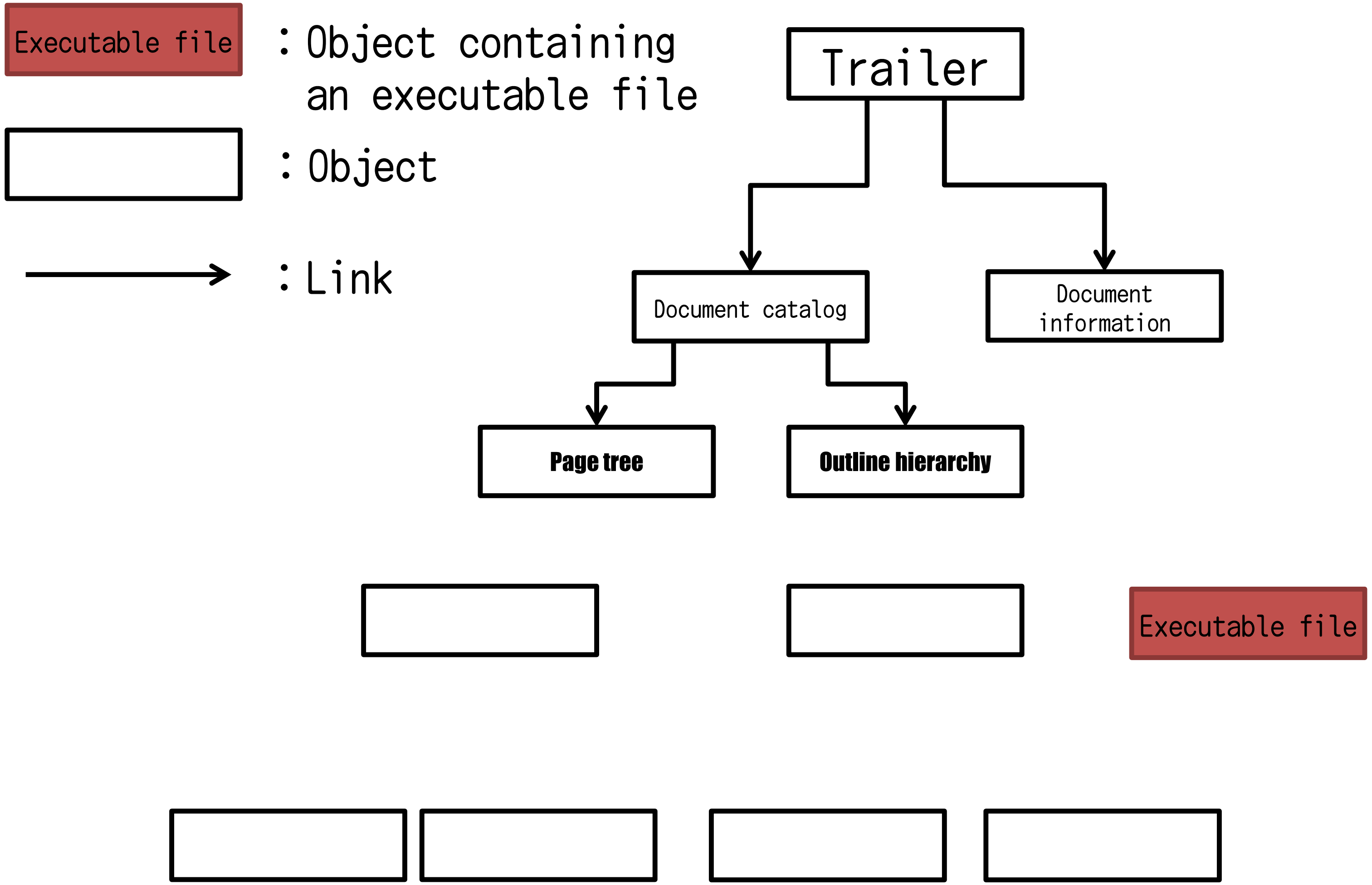


A PDF file containing an executable file

When an executable file is inserted as an object in disregard of document structure, it is often unreferenced.

AS7:Unreferenced object

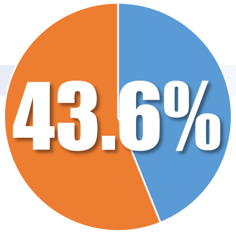


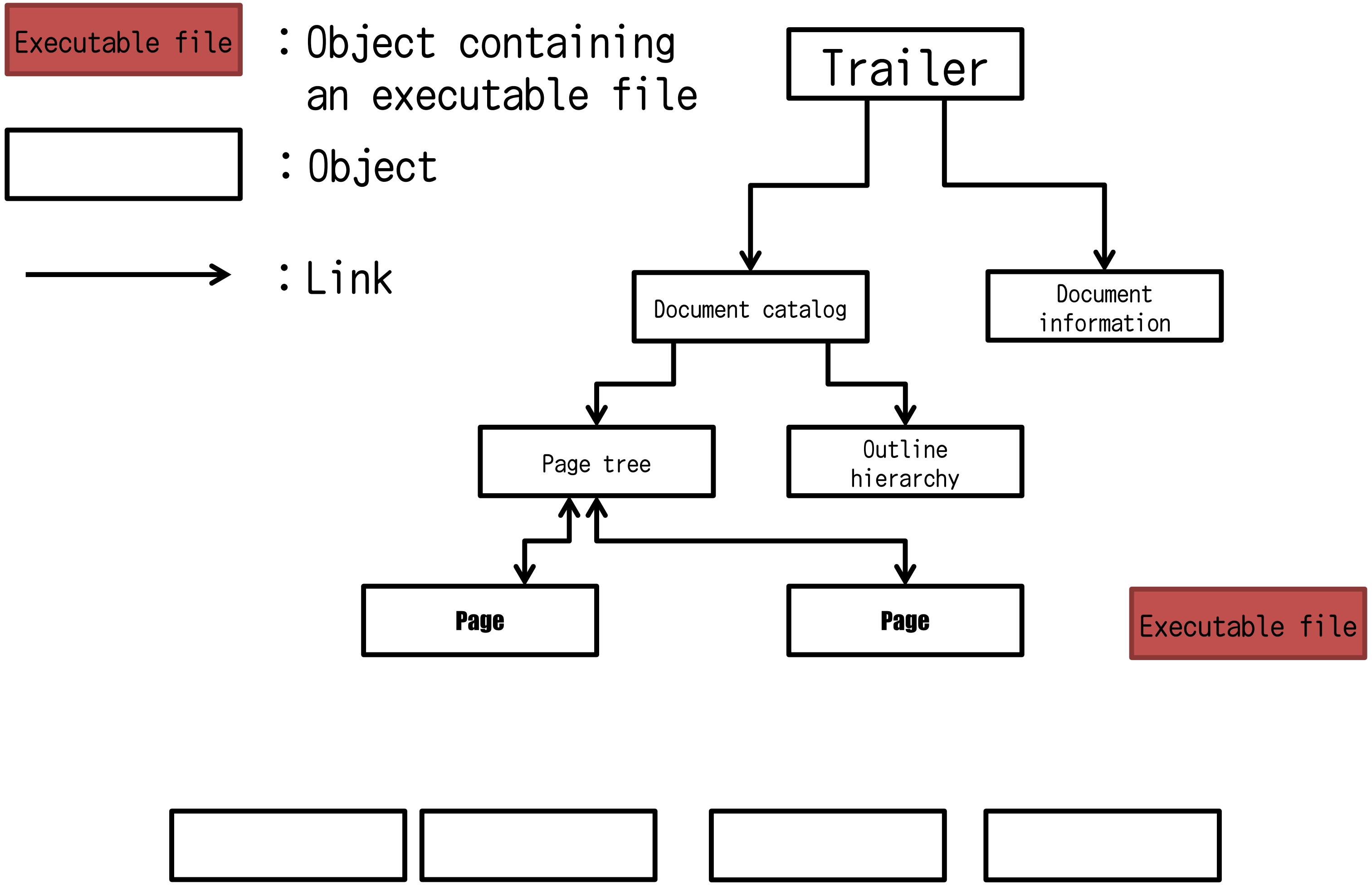


A PDF file containing an executable file

When an executable file is inserted as an object in disregard of document structure, it is often unreferenced.

AS7:Unreferenced object





Executable file ：Object containing an executable file

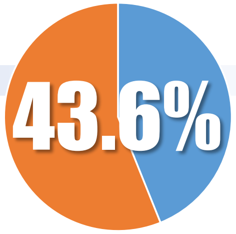
：Object

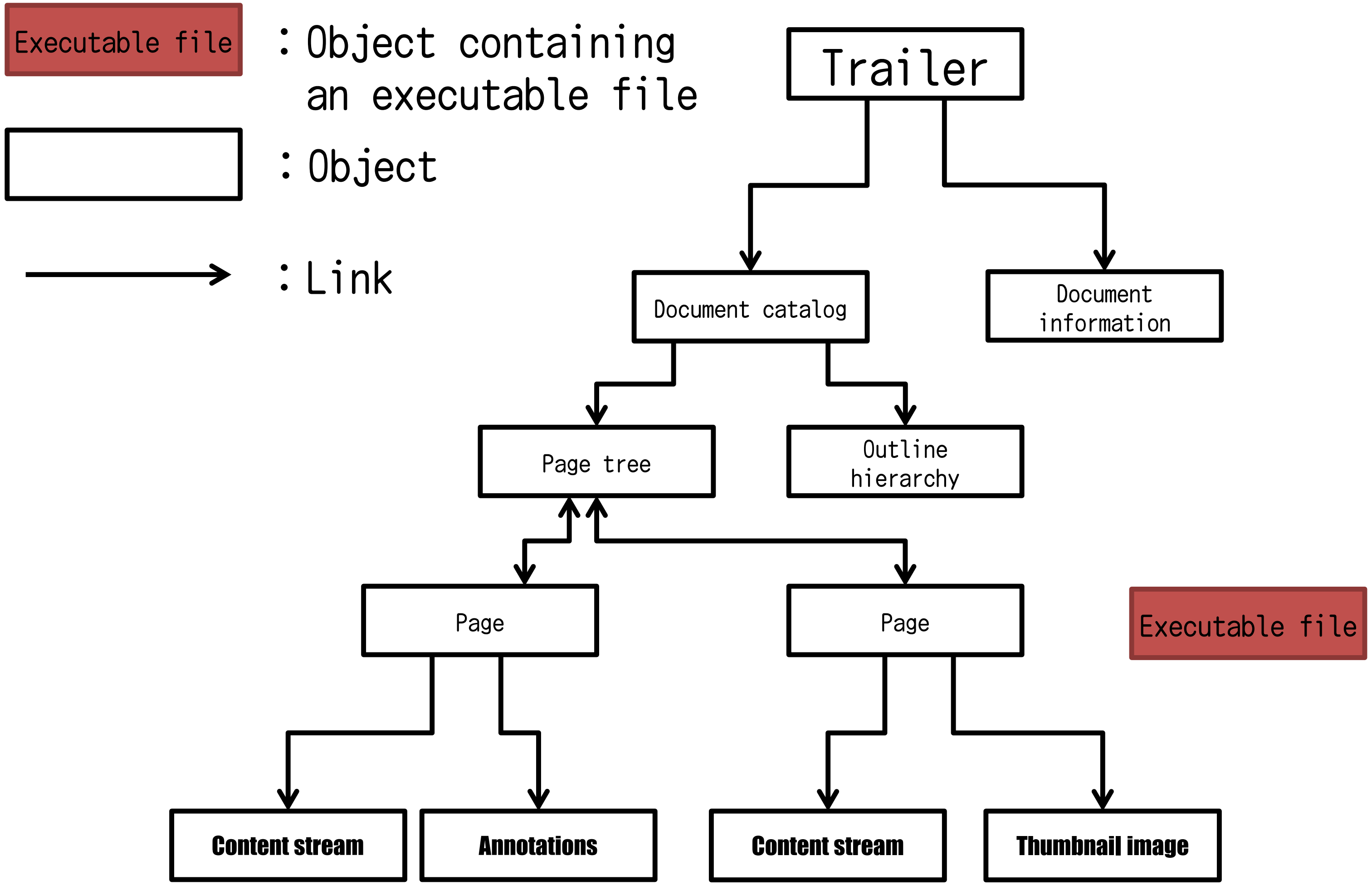
：Link

A PDF file containing an executable file

When an executable file is inserted as an object in disregard of document structure, it is often unreferenced.

AS7:Unreferenced object

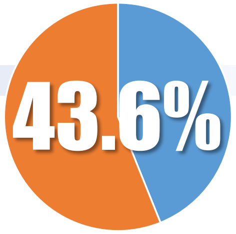


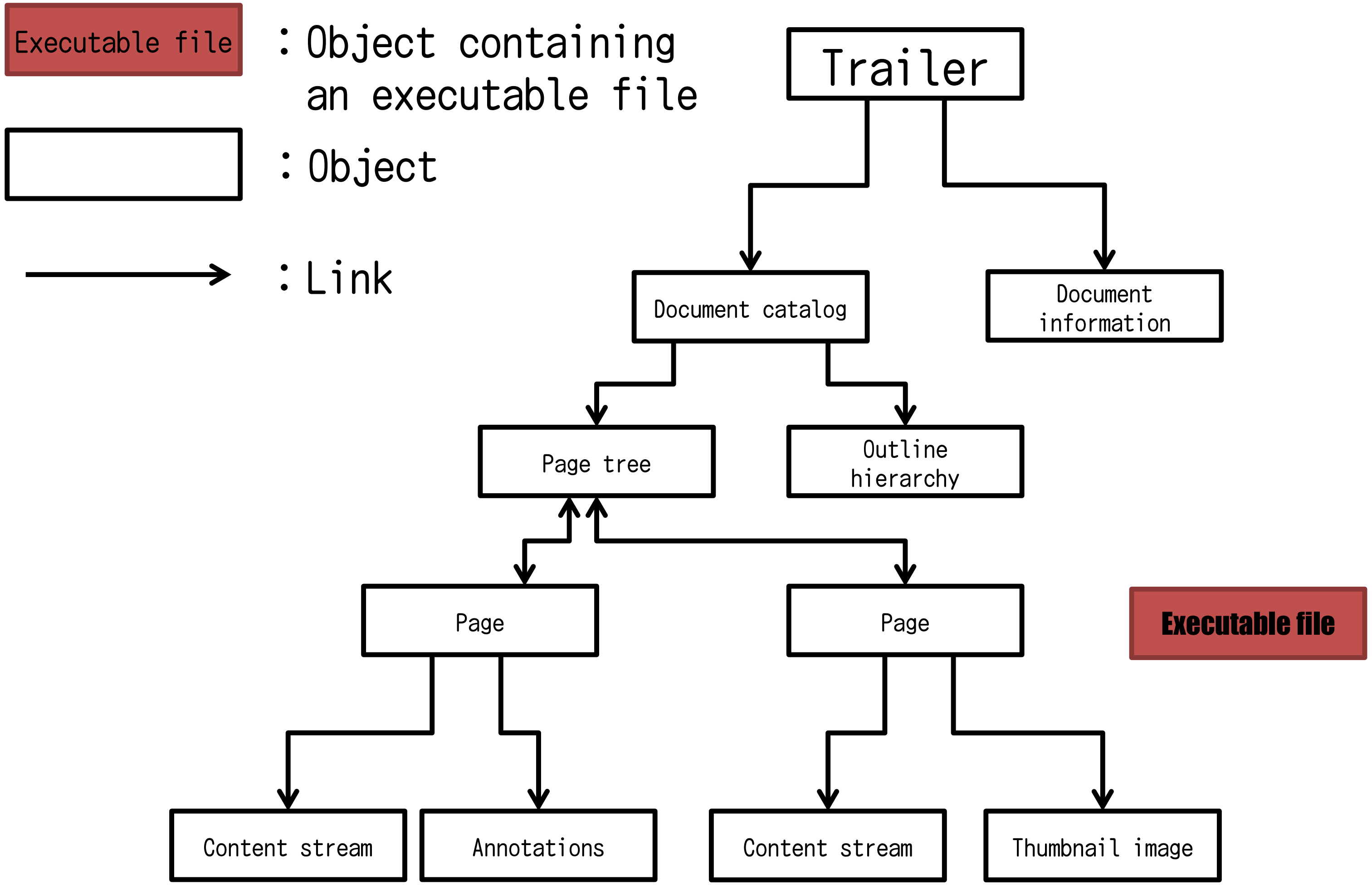


A PDF file containing an executable file

When an executable file is inserted as an object in disregard of document structure, it is often unreferenced.

AS7:Unreferenced object





🟥 Executable file : Object containing an executable file

□ : Object

→ : Link

Trailer

Document catalog

Document information

Page tree

Outline hierarchy

Page

Page

Executable file

Content stream

Annotations

Content stream

Thumbnail image

A PDF file containing an executable file

When an executable file is inserted as an object in disregard of document structure, it is often unreferenced.

How to run “o-checker”

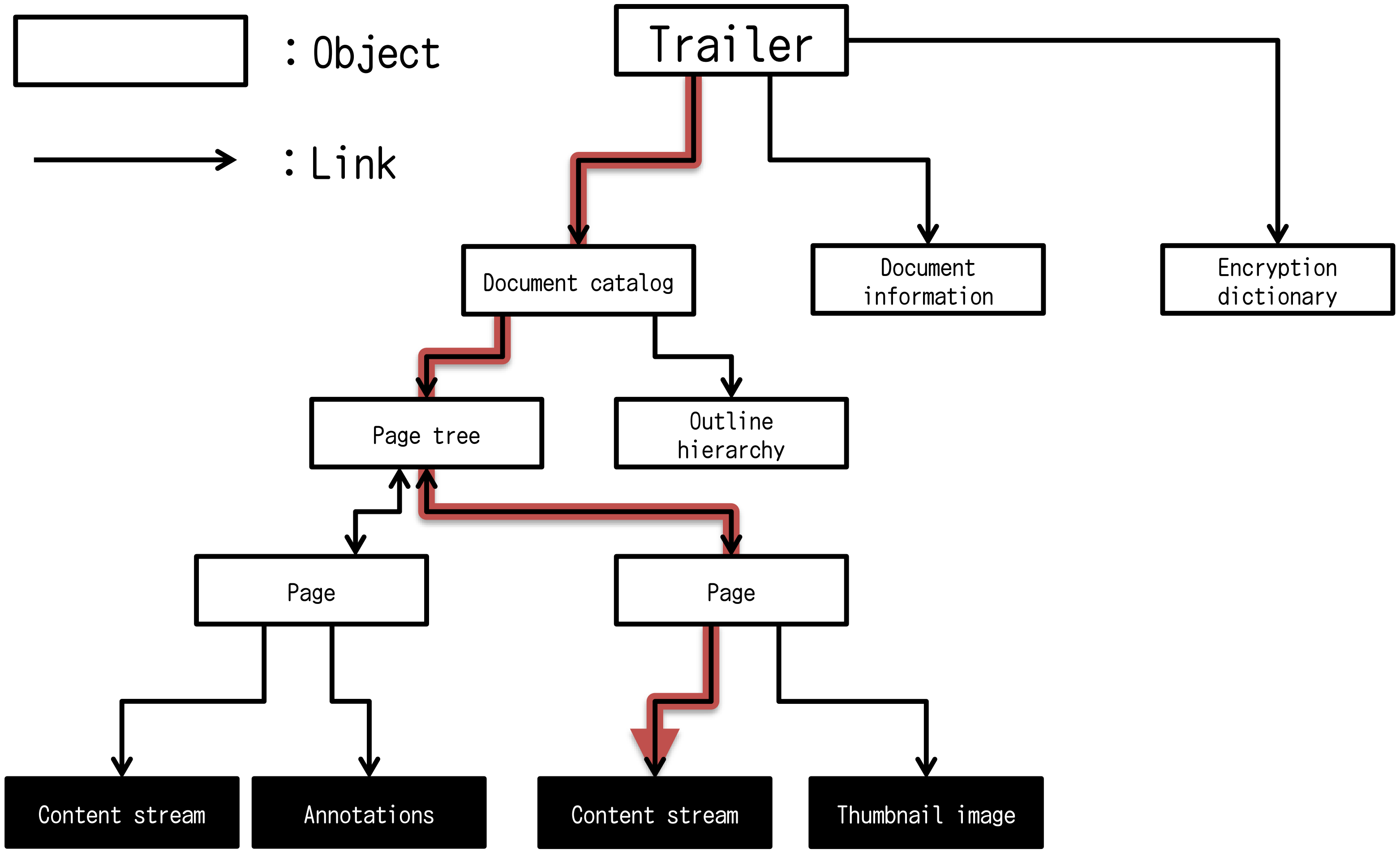
- Requirement
  - Python 2.7.3 or later
  - Any OSes that can run Python
  - PyCrypto for 2.7
    (for an encrypted PDF file)

- [command example]

```
> python o-checker.py malware.doc
```

DEMO

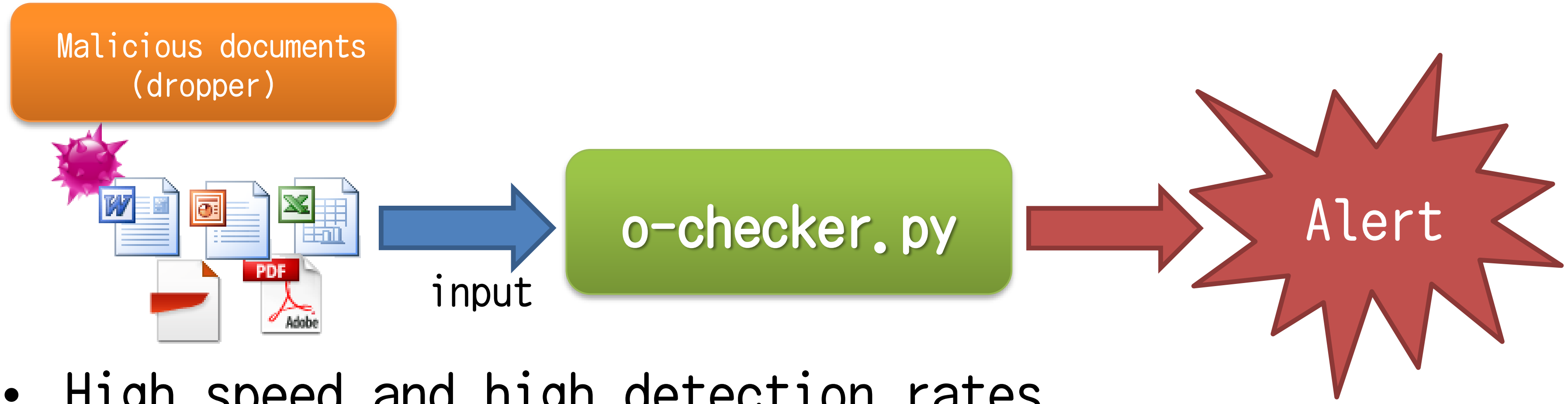
Structure of PDF：Encryption



Structure of a PDF document enctypted

Encryption applies to almost all strings and streams in the PDF file. Leaving the other object types unencrypted allows random access to the objects within a document. (except for the object stored in ObjStm)

Conclusion



- High speed and high detection rates
- Almost maintenance-free
- MIT License

  Available from
  Black Hat USA 2016 web site

Thank you!