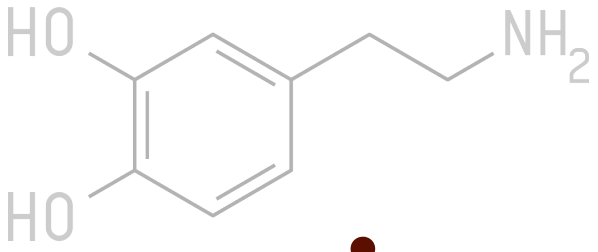


Schizophrenic files

AREA41

Ange Albertini
Gynvael Coldwind

2014/06/02
Zürich, Switzerland

Gynvael Coldwind

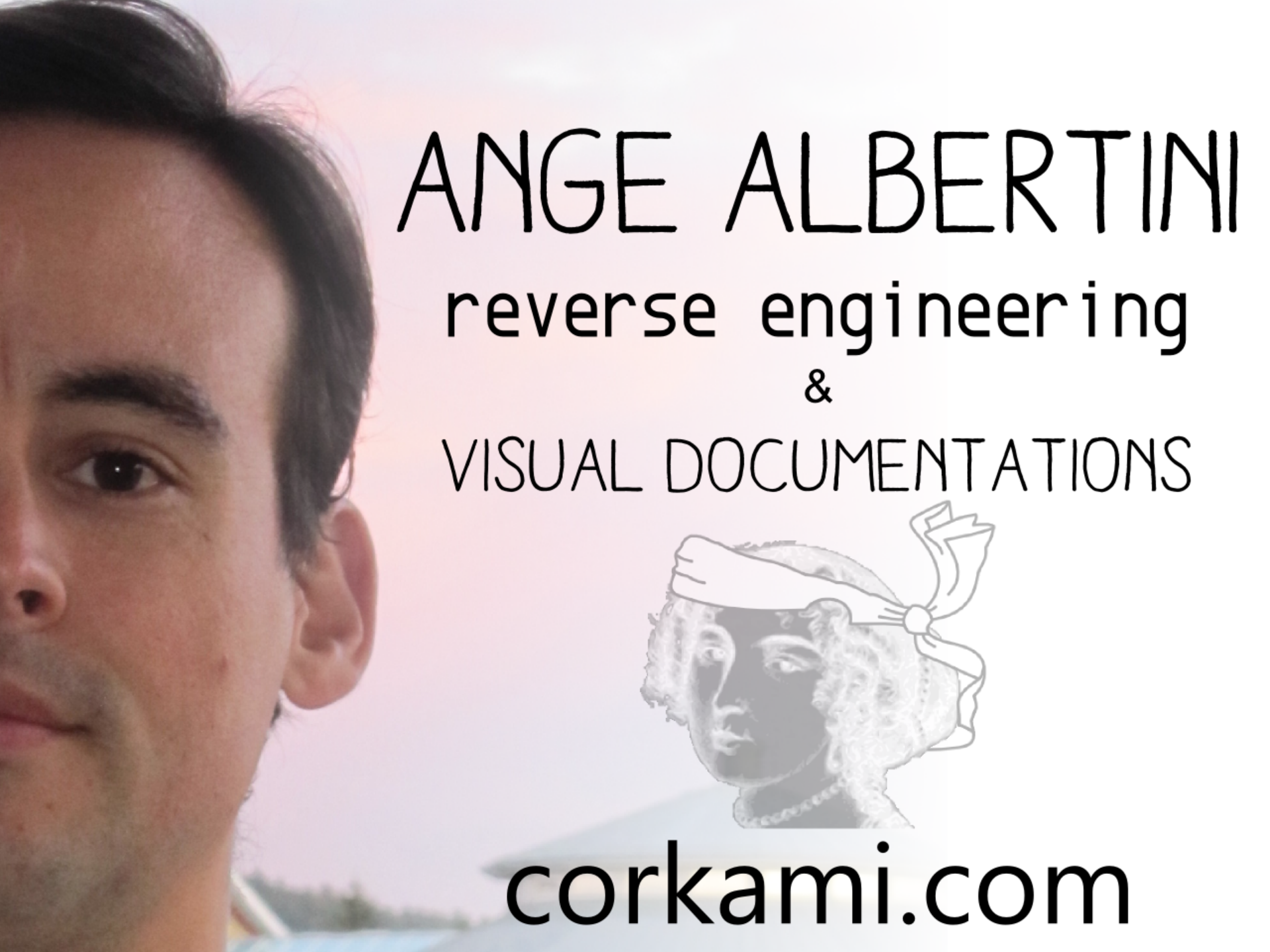
Security researcher, Google

Dragon Sector captain

Likes hamburgers

http://gynvael.coldwind.pl/

All opinions expressed during this presentation are mine and mine alone.
They are not opinions of my lawyer, barber and especially not my employer.


ANGE ALBERTINI
reverse engineering
&
VISUAL DOCUMENTATIONS
corkami.com

1 file + 2 tools
⇒ 2 different documents

No active detection in the file.

abusing parsers for

- fun
- bypassing security
  - same-origin policy
  - evade detection
  - exfiltration
  - signing
    - Android Master Key

excerpt from Gynvael's talk:

"Dziesięć tysięcy pułapek: ZIP, RAR, etc."

(http://gynvael.coldwind.pl/?id=523)

ZIP

trick 1

a glitch in the matrix

file names in ZIP

a couple of files with the same name?

update:

for an awesome example see:

Android: One Root to Own Them All

Jeff Forristal / Bluebox

(https://media.blackhat.com/us-13/US-13-Forristal-Android-One-Root-to-Own-Them-All-Slides.pdf)

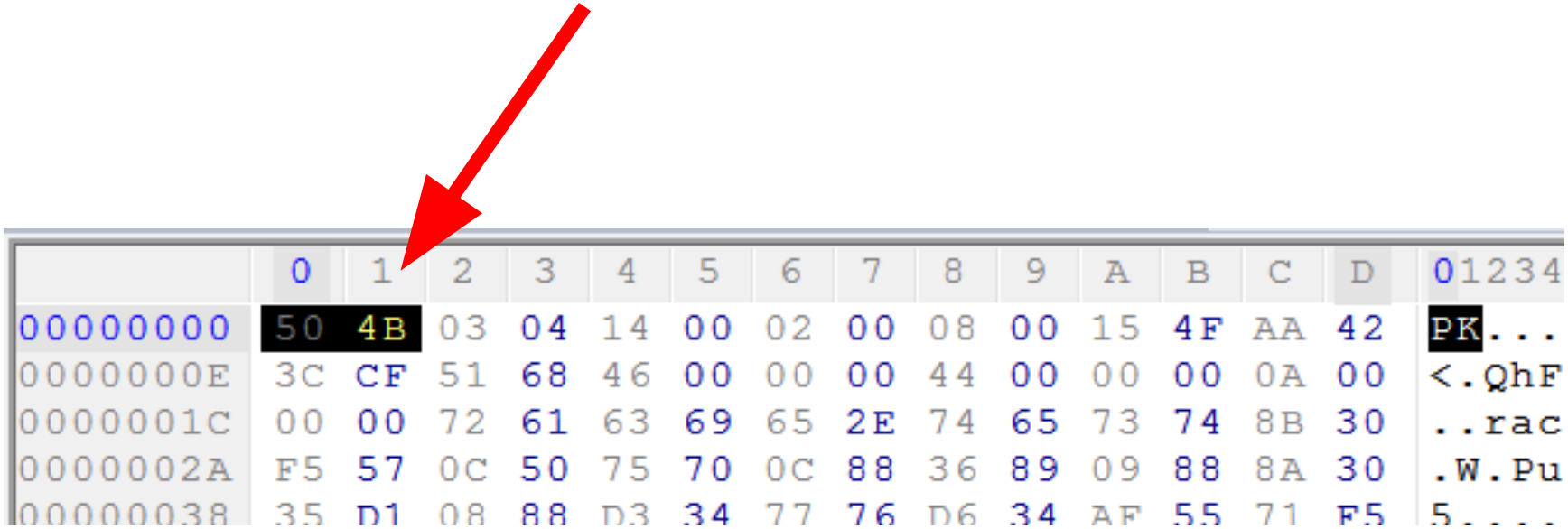
ZIP

trick 2

abstract kitty

Let's start with simple stuff - the ZIP format

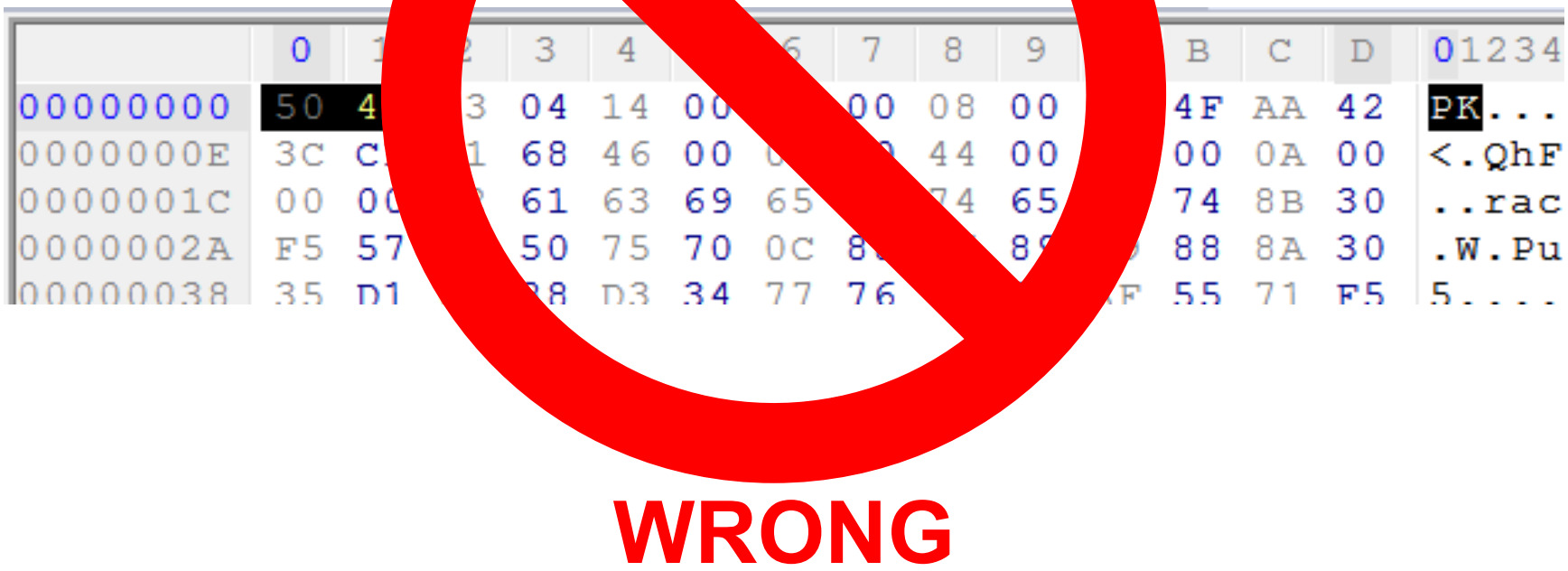
A ZIP file begins with letters **PK.**



	0	1	2	3	4	5	6	7	8	9	A	B	C	D	01234
00000000	50	4B	03	04	14	00	02	00	08	00	15	4F	AA	42	PK...
0000000E	3C	CF	51	68	46	00	00	00	44	00	00	00	0A	00	<.QhF
0000001C	00	00	72	61	63	69	65	2E	74	65	73	74	8B	30	..rac
0000002A	F5	57	0C	50	75	70	0C	88	36	89	09	88	8A	30	.W.Pu
00000038	35	D1	08	88	D3	34	77	76	D6	34	AF	55	71	F5	5....

Let's start with simple stuff - the ZIP format

A ZIP file begins with letters **PK.**



WRONG

ZIP - second attempt :)

last 65557 bytes of the file
the "header" is
"somewhere" here

.zip file

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	0123456789ABCD
00000000	50	4B	03	04	14	00	02	00	08	00	15	4F	AA	42	PK.....O.B
0000000E	3C	CF	51	68	46	00	00	00	44	00	00	00	0A	00	<.QhF...D.....
0000001C	00	00	72	61	63	69	65	2E	74	65	73	74	8B	30	..racie.test.0
0000002A	F5	57	0C	50	75	70	0C	88	36	89	09	88	8A	30	.W.Pup..6....0
00000038	35	D1	08	88	D3	34	77	76	D6	34	AF	55	71	F5	5....4wv.4.Uq.
00000046	74	76	0C	D2	0D	0E	71	F4	73	71	0C	72	D1	75	tv....q.sq.r.u
00000054	F4	0B	F1	0C	F3	0C	0A	0D	D6	0D	71	0D	0E	D1	..........q...
00000062	75	F3	F4	71	55	54	F1	D0	F6	D0	02	00	50	4B	u..qUT......PK
00000070	01	02	14	00	14	00	02	00	08	00	15	4F	AA	42	..........O.B
0000007E	3C	CF	51	68	46	00	00	00	44	00	00	00	0A	00	<.QhF...D.....
0000008C	00	00	00	00	00	00	01	00	20	00	00	00	00	00	....
0000009A	00	00	72	61	63	69	65	2E	74	65	73	74	50	4B	..racie.testPK
000000A8	05	06	00	00	00	00	01	00	01	00	38	00	00	00	..........8...
000000B6	6E	00	00	00	00	00									n.....

ZIP - "somewhere" ?!

you begin ZIP parsing from this; it MUST be at the end of the file

4.3.16 End of central directory record:

22 bajty

end of central dir signature	4 bytes	(0x06054b50)
number of this disk	2 bytes	
number of the disk with the start of the central directory	2 bytes	
total number of entries in the central directory on this disk	2 bytes	
total number of entries in the central directory	2 bytes	
size of the central directory	4 bytes	
offset of start of central directory with respect to the starting disk number	4 bytes	
.ZIP file comment length	**2 bytes**	
.ZIP file comment	**(variable size)**	

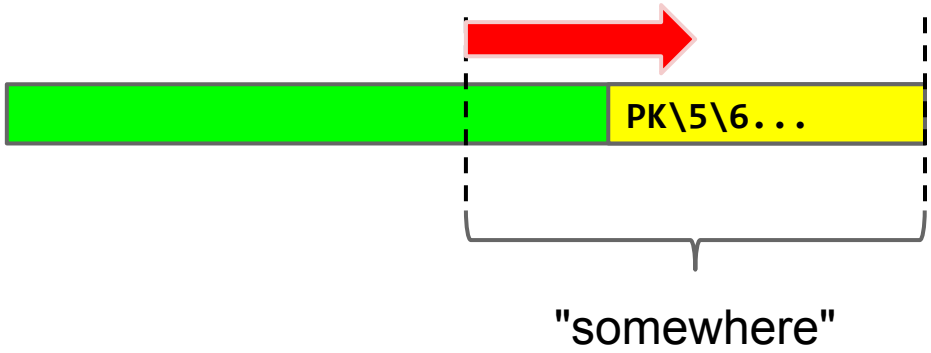
$0000-$FFFF
0-65535

Total: from 22 to 65557 bytes

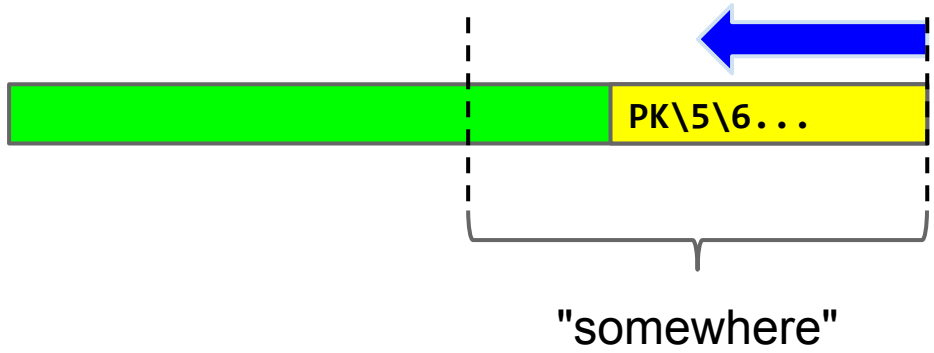
(aka: PK\5\6 magic will be somewhere between EOF-65557 and EOF-22)

ZIP - looking for the "header"?

"From the START"
Begin at **`EOF-65557`**,
and move forward.



"From the END"
(ZIPs usually don't have comments)
Begin at **`EOF-22`**,
and move backward.



The show will continue in a moment.

Larch

Something completely different

ZIP Format - LFH

4.3.7 Local file header:

random stuff

local file header signature	4 bytes	(0x04034b50)
version needed to extract	2 bytes	
general purpose bit flag	2 bytes	
compression method	2 bytes	
last mod file time	2 bytes	
last mod file date	2 bytes	
crc-32	4 bytes	
compressed size	4 bytes	
uncompressed size	4 bytes	
file name length	2 bytes	
extra field length	2 bytes	

file name (variable size)
extra field (variable size)
file data (variable size)

PK\3\4... LFH + data

Each file/directory in a ZIP has LFH + data.

ZIP Format - CDH

[central directory header n]

central file header signature	4 bytes	(0x02014b50)
version made by	2 bytes	
version needed to extract	2 bytes	
general purpose bit flag	2 bytes	
compression method	2 bytes	
last mod file time	2 bytes	
last mod file date	2 bytes	
crc-32	4 bytes	
compressed size	4 bytes	
uncompressed size	4 bytes	
file name length	2 bytes	
extra field length	2 bytes	
file comment length	2 bytes	
disk number start	2 bytes	
internal file attributes	2 bytes	
external file attributes	4 bytes	
relative offset of local header	4 bytes	
file name (variable size)		
extra field (variable size)		
file comment (variable size)		

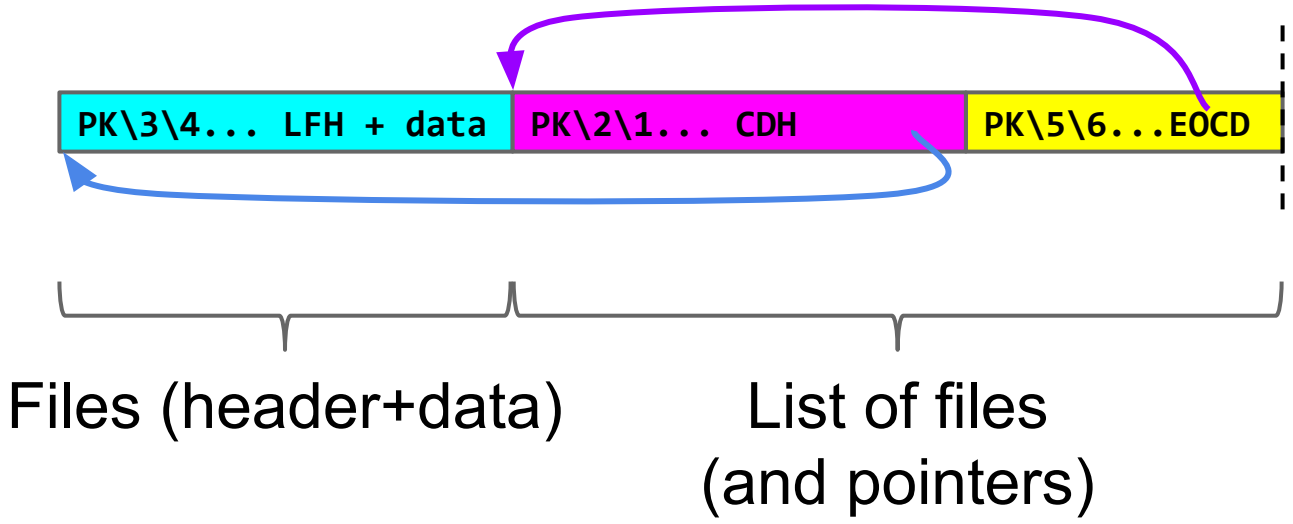
similar stuff to LFH

thanks to the redundancy you can recover LFH using CDH, or CDH using LFH

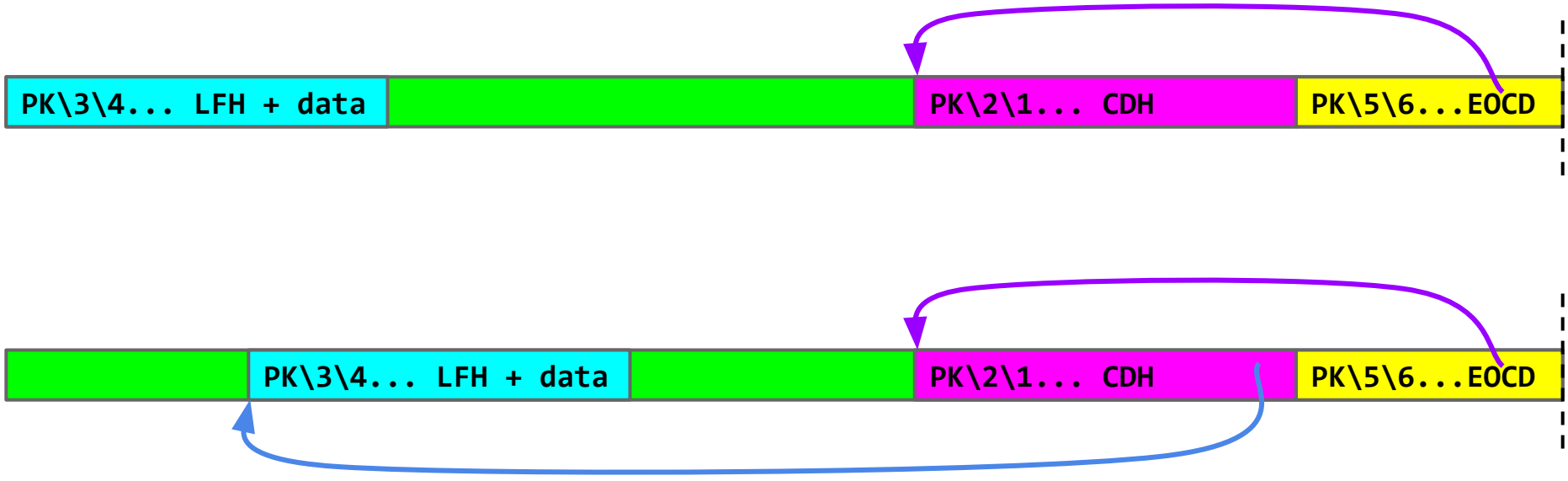
PK\2\1... CDH

Each file/directory has a CDH entry in the Central Directory

ZIP - a complete file

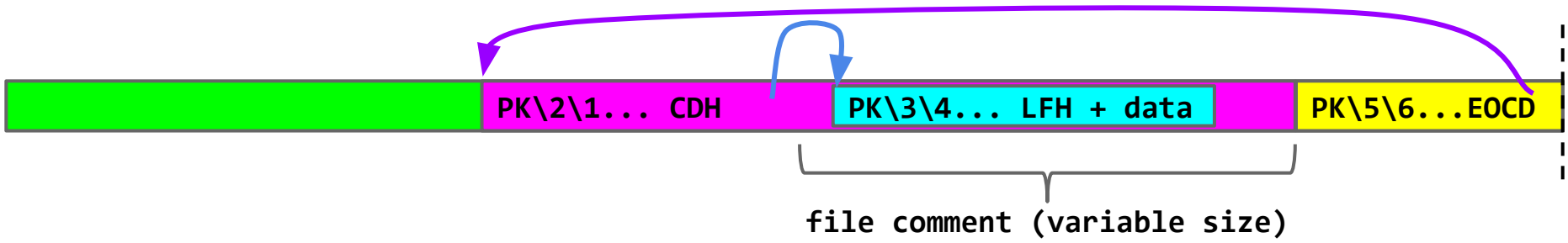


ZIP - a complete file (continued)



If the list of the files has pointers to files...
... the ZIP structure can be more relaxed.

ZIP - a complete file (continued)



You can even do an "inception"
(some parsers may allow **EOCD(CHD(LFH))**)

And now back to our show!

(we were looking for the EOCD)

Larch

Something completely different

ZIP - looking for the "header"?

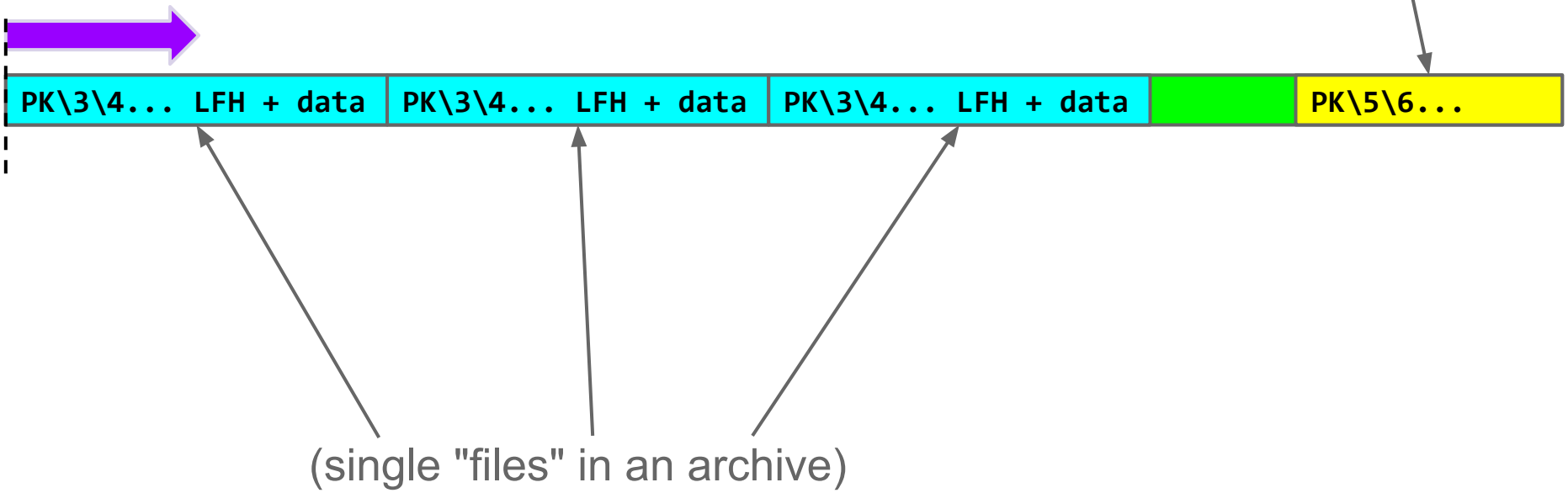
(who cares...)

"stream"

Let's ignore EOCD!

(it's sometimes faster)

(99.9% of ZIPs out there can be parsed this way)



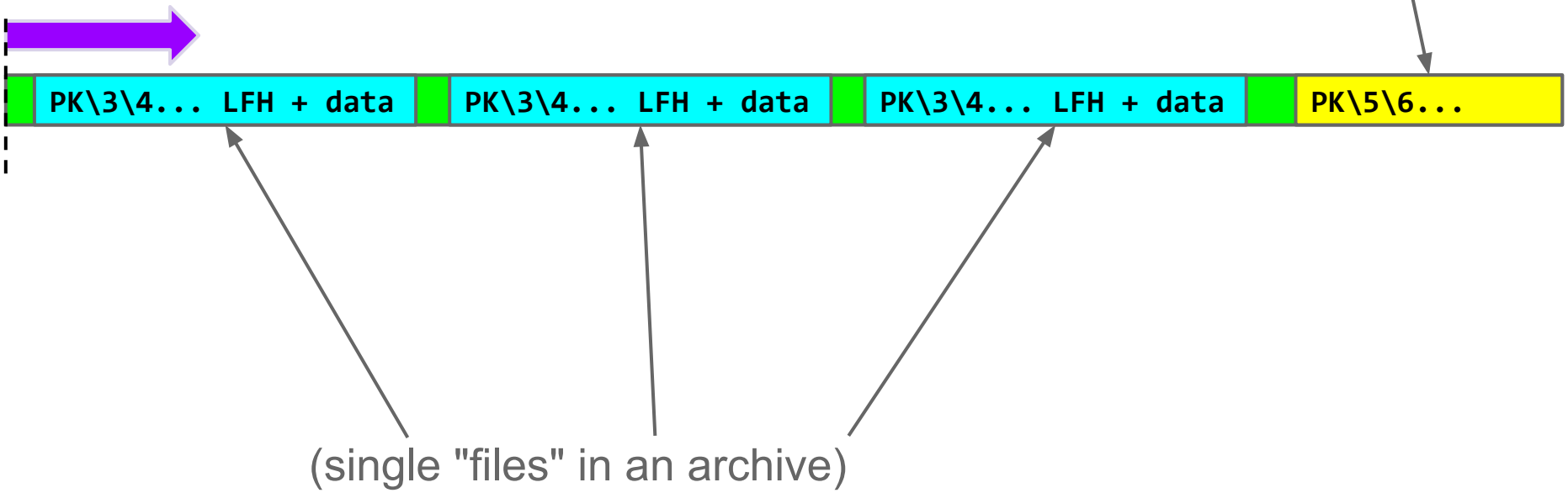
ZIP - looking for the "header"?

(who cares...)

"aggressive stream"

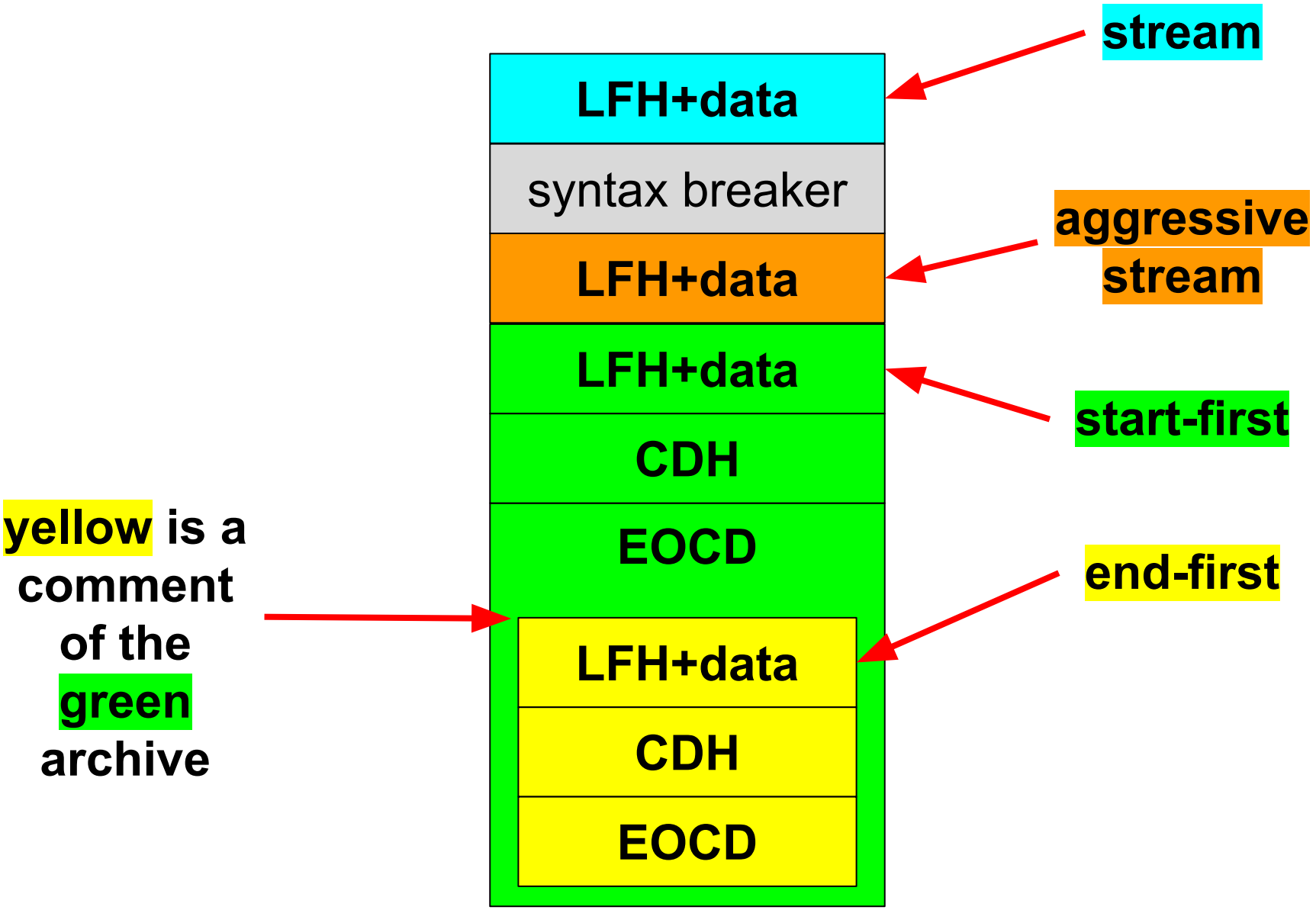
We ignore the "garbage"!

(forensics)



Let's test the parsers!
abstract.zip

abstract.zip



abstract.zip

```
from zipfile import ZipFile
ZipFile("abstract.zip", "r").
printdir()
```

```
gynvael:haven-windows> zip.py
File Name
readme_EndFirst.txt
```

abstract.zip

```php
<?php
$za = new ZipArchive();
$za->open('abstract.zip');
for ($i=0; $i<$za->numFiles;$i++) {
    echo "index: $i\n";
    print_r($za->statIndex($i));
}
echo "numFile:" . $za->numFiles . "\n";
```

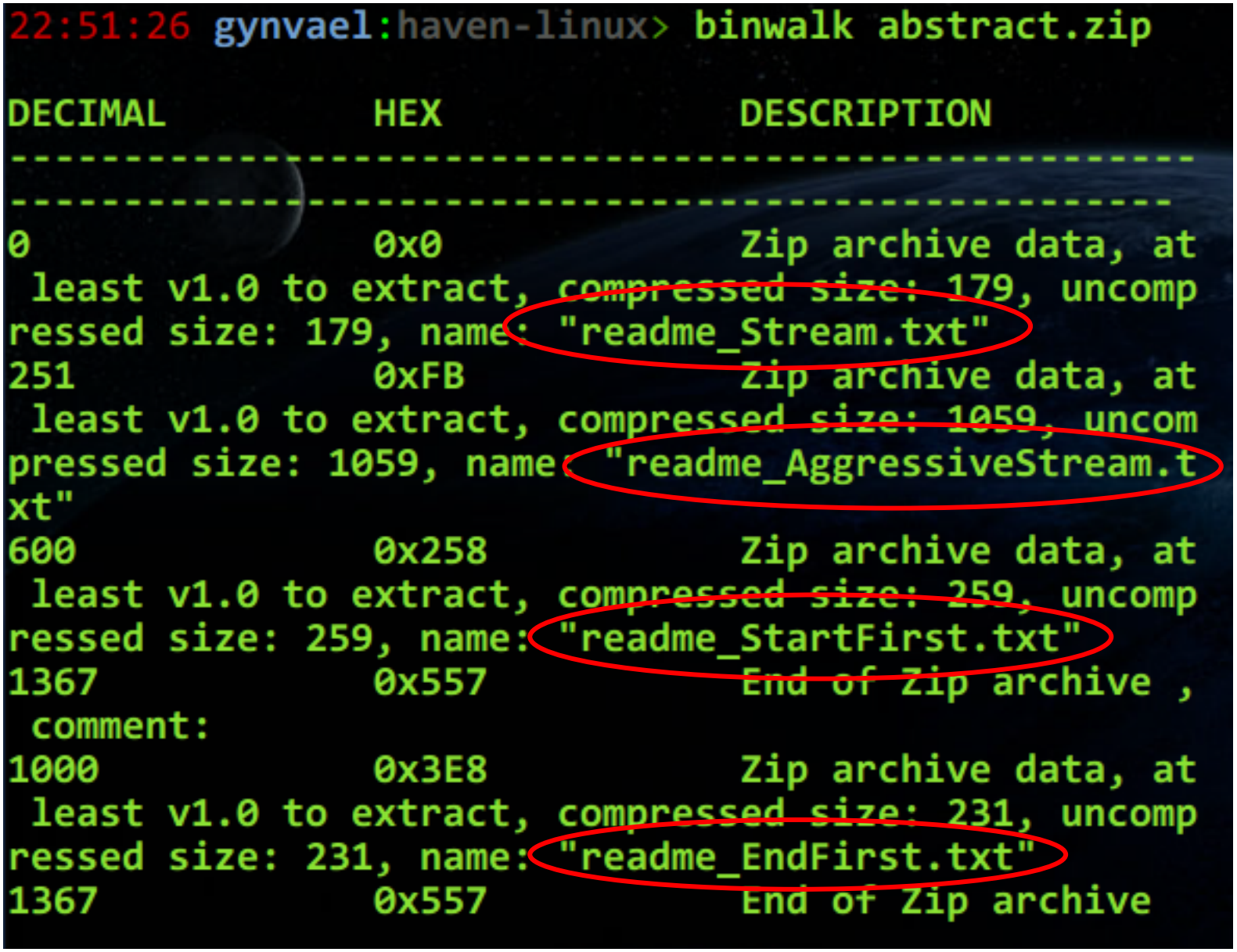
```
gynvael:haven-windows> php zip.php
index: 0
Array
(
    [name] => readme_StartFirst.txt
    [index] => 0
    [crc] => 543868170
    [size] => 259
    [mtime] => 312764400
    [comp_size] => 259
    [comp_method] => 0
)
numFile:1
```

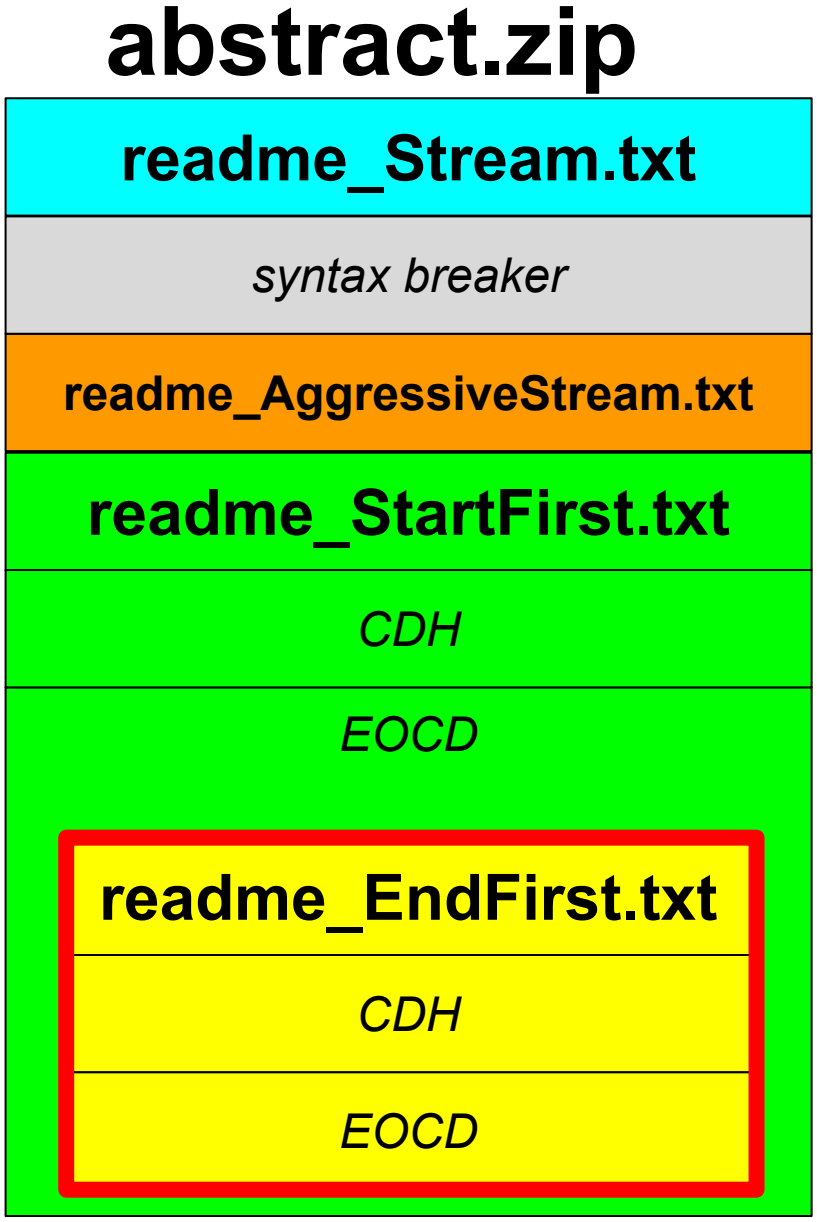
abstract.zip

```
import java.io.FileInputStream;
import java.io.InputStream;
import java.util.zip.ZipEntry;
import java.util.zip.ZipInputStream;

public class zip {
  public static void main(String args[]) throws
       java.io.IOException, java.io.FileNotFoundException {
    InputStream f = new FileInputStream("abstract.zip");
    ZipInputStream z = new ZipInputStream(f);

    ZipEntry e;
    while((e = z.getNextEntry()) != null) {
      System.out.println(e.getName());
    }
  }
}
```

```
gynvael:haven-windows> java zip
readme_Stream.txt
```

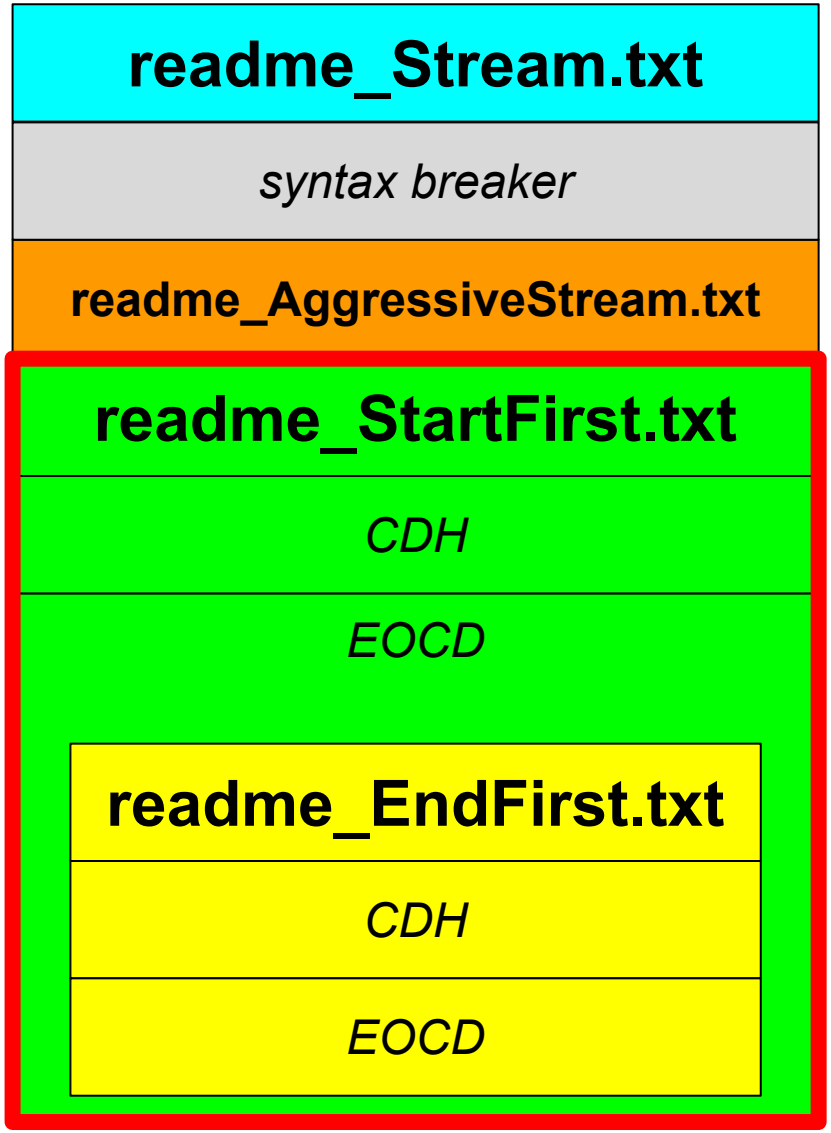
abstract.zip



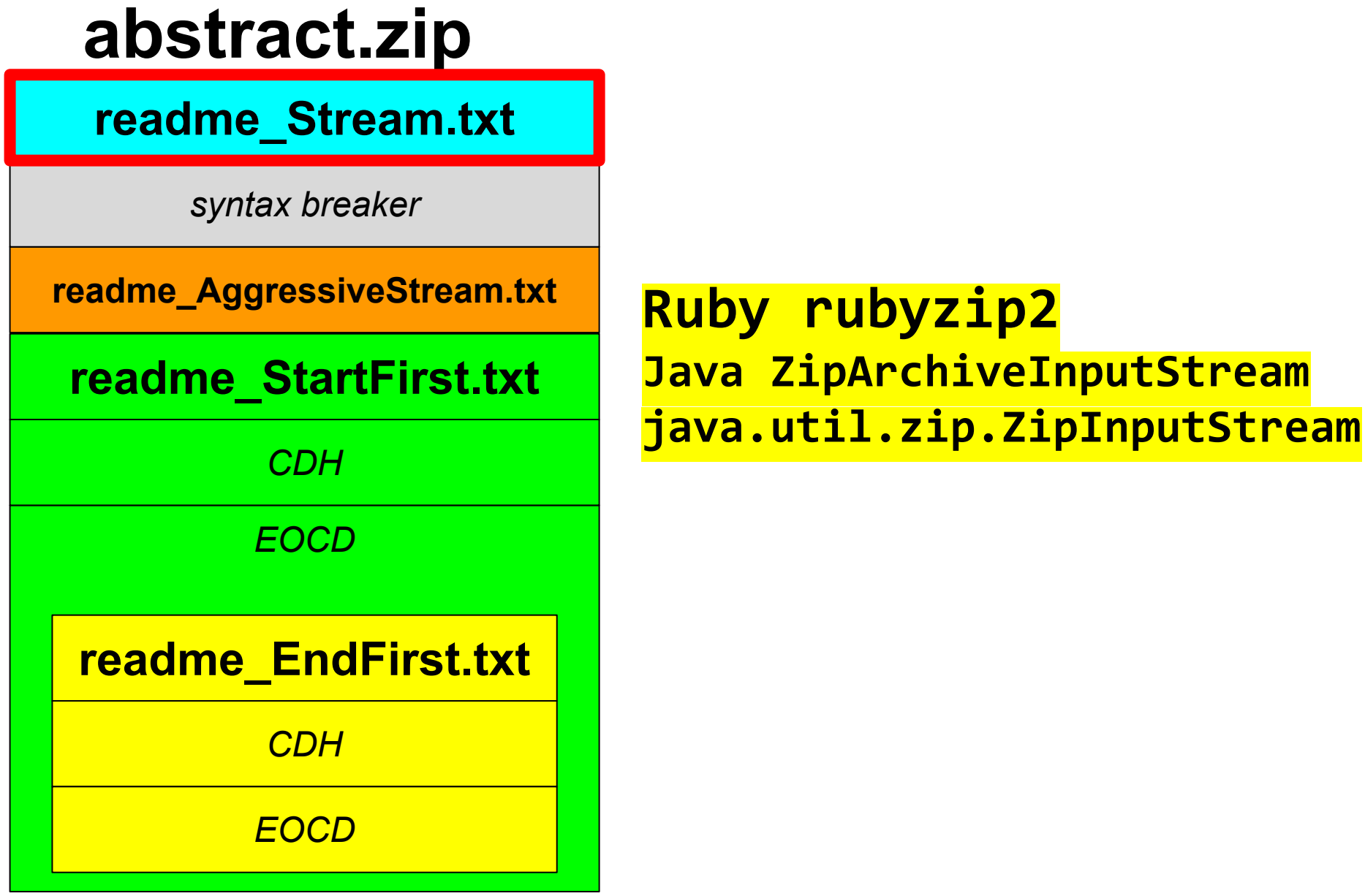


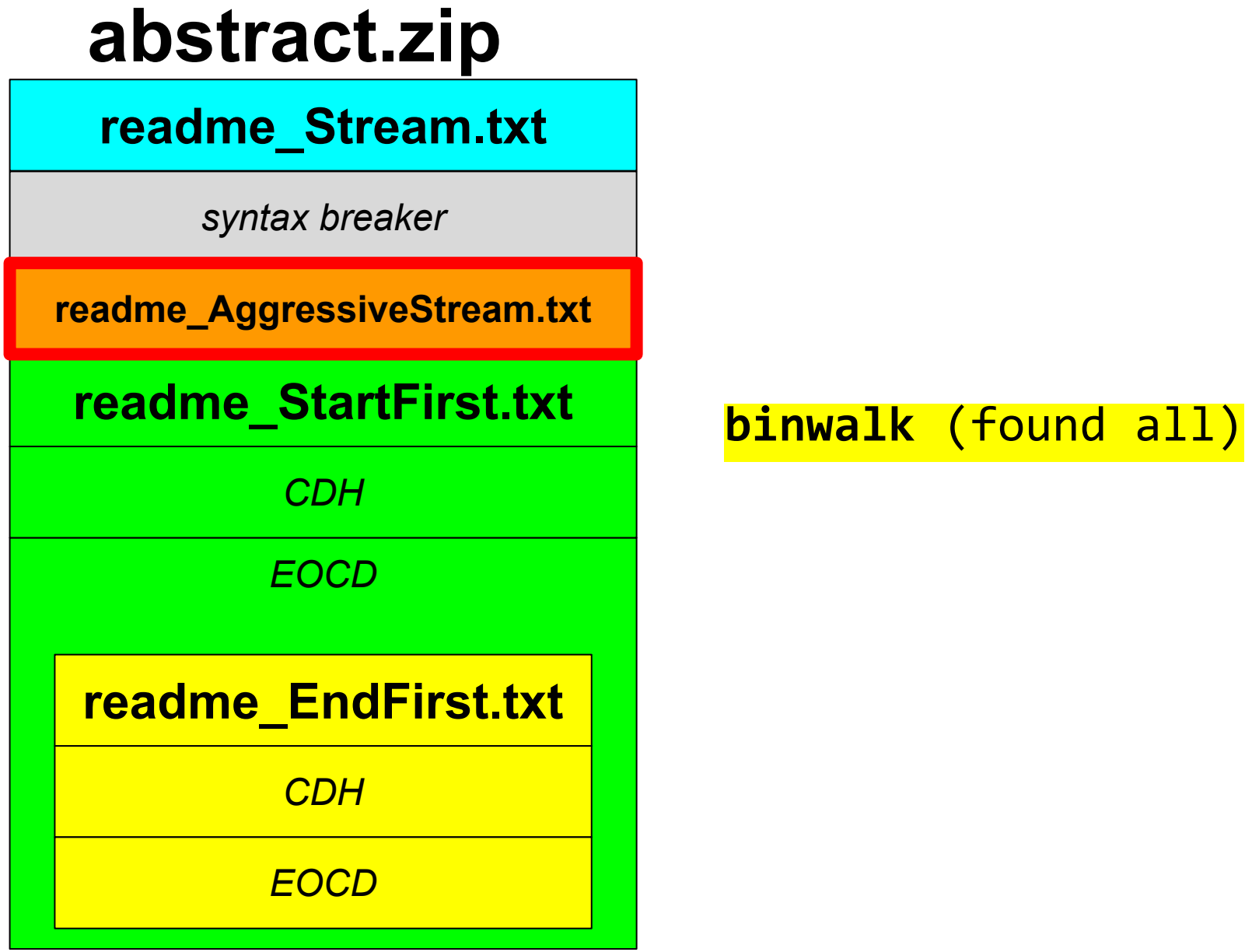
Total Commander 8.01
UnZip 6.00 (Debian)
Midnight Commander
Windows 7 Explorer
ALZip
KGB Archiver
7-zip
b1.org
Python zipfile
JSZip
C# DotNetZip
perl Archive::Zip
Jeffrey's Exif Viewer
WOBZIP
GNOME File Roller
WinRAR
OSX UnZip
zip.vim v25
Emacs Zip-Archive mode
Ada Zip-Ada v45
Go archive/zip
Pharo smalltalk 2.0 ZipArchive
Ubuntu less
Java ZipFile

abstract.zip



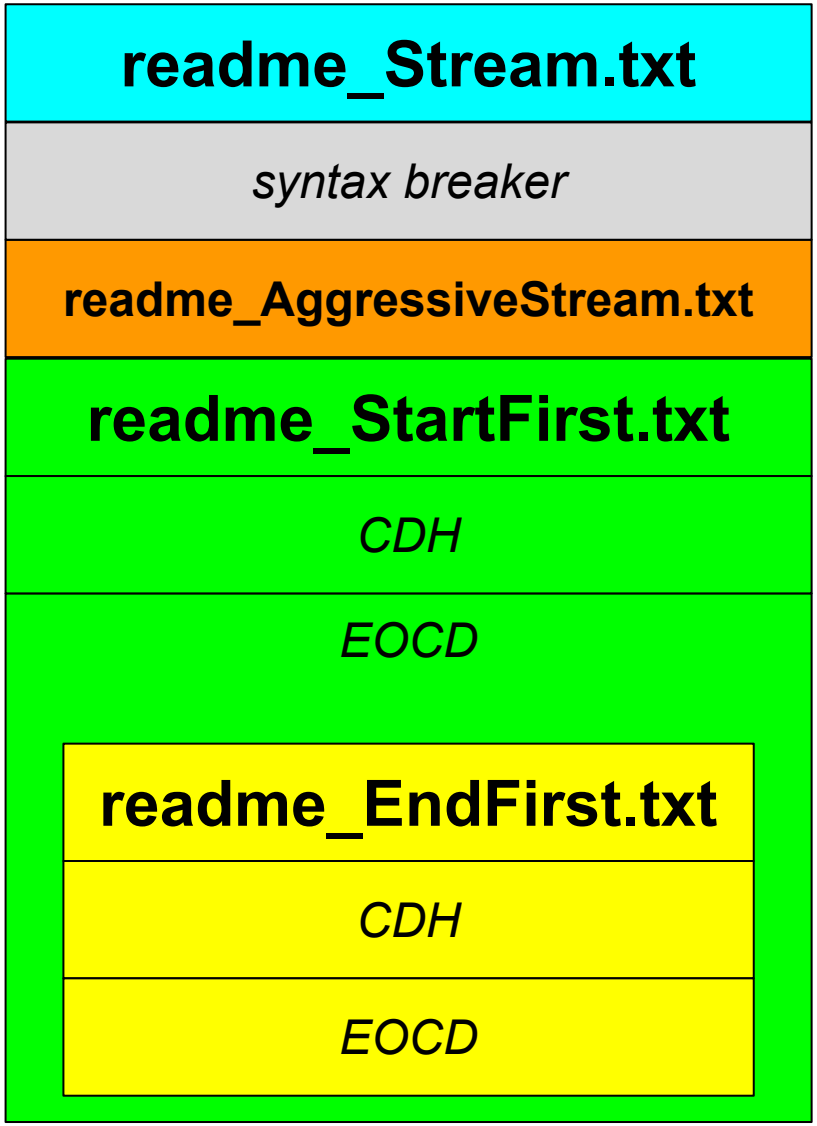
PHP ZipArchive
PHP zip_open ...
PHP zip:// wrapper
tcl + tclvfs + tclunzip


abstract.zip
readme_Stream.txt
syntax breaker
readme_AggressiveStream.txt
readme_StartFirst.txt
CDH
EOCD
readme_EndFirst.txt
CDH
EOCD
Ruby rubyzip2
Java ZipArchiveInputStream
java.util.zip.ZipInputStream



binwalk (found all)

abstract.zip - result summary



Thanks!

- Mulander
- Felix Groebert
- Salvation
- j00ru

abstract.zip - who cares?

- verify files via End-First
- unpack via Stream

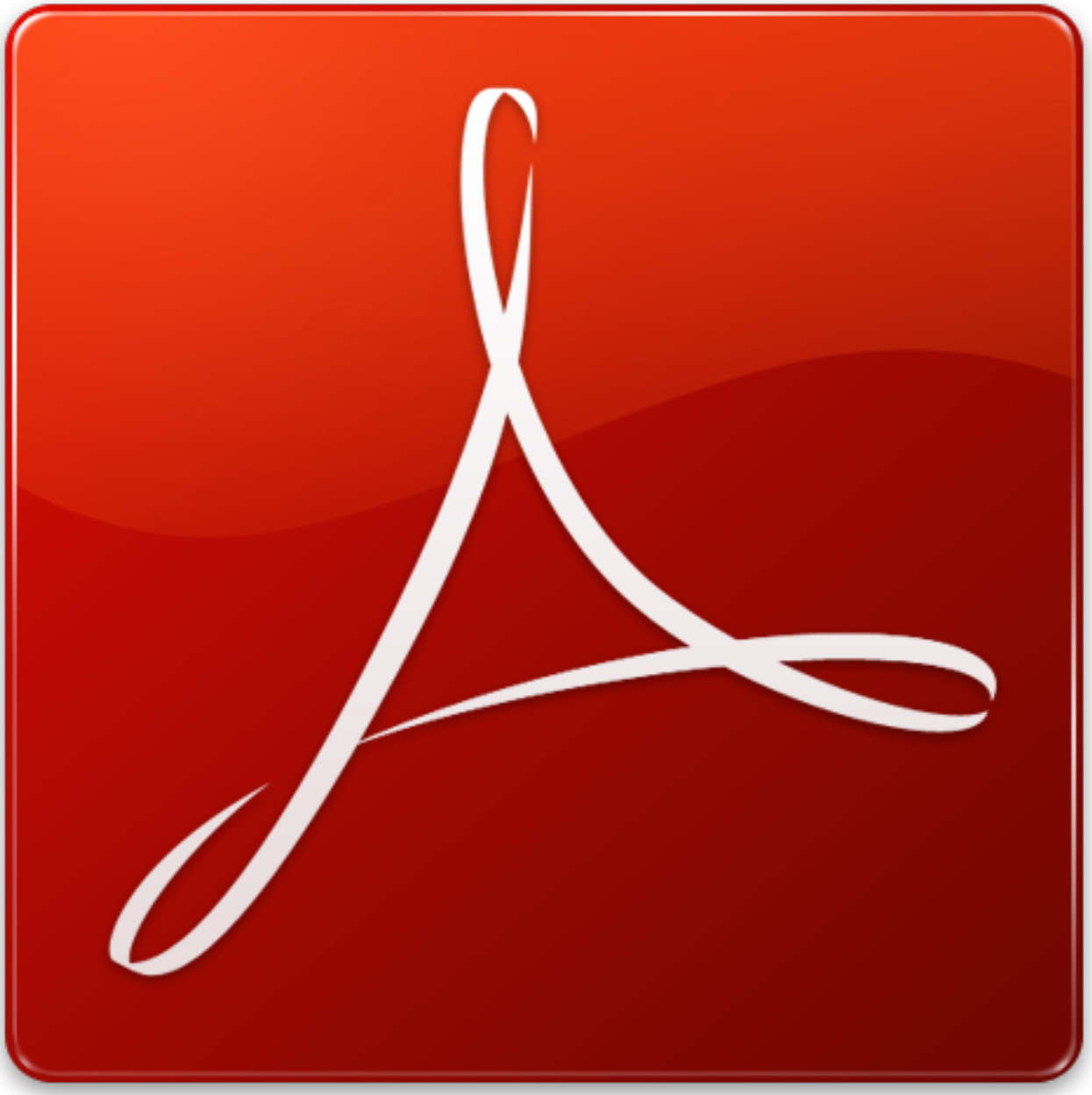
Oops.

abstract.zip - AV

EICAR test results (using VT):

- most End-First
- some Aggressive
- Stream-only:
  - VBA32
  - NANO-Antivirus
  - Norman
  - F-Prot
  - Agnitum
  - Commtouch

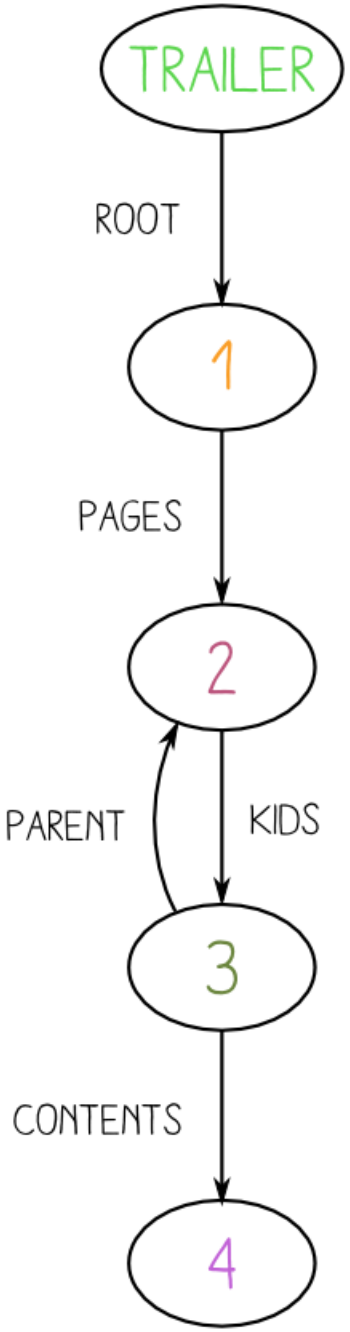
https://docs.google.com/spreadsheet/ccc?key=0Apy5AGVPzpIOdDRPTFNJQXpqNkdjUzl4SE80c1kwdkE&usp=sharing



http://youtu.be/JQrBgVRgqtc?t=11m15s
https://speakerdeck.com/ange/pdf-secrets-hiding-and-revealing-secrets-in-pdf-documents?slide=44


```
%PDF-1.1

1 0 obj
<<
  /Pages 2 0 R
>>
endobj

2 0 obj
<<
  /Type /Pages
  /Count 1
  /Kids [3 0 R]
>>
endobj

3 0 obj
<<
  /Type /Page
  /Contents 4 0 R
  /Parent 2 0 R
  /Resources <<
    /Font <<
      /F1 <<
        /Type /Font
        /Subtype /Type1
        /BaseFont /Arial
      >>
    >>
  >>
>>
endobj
```

```
4 0 obj
<< /Length 50 >>
stream
BT
  /F1 110 Tf
  10 400 Td
  (Hello World!)Tj
ET
endstream
endobj

xref
0 5
0000000000 65535 f
0000000010 00000 n
0000000047 00000 n
0000000111 00000 n
0000000313 00000 n

trailer
<<
  /Root 1 0 R
>>

startxref
413
%%EOF
```



7.5.5 File Trailer

The *trailer* of a PDF file enables a conforming reader to quickly find the cross-reference table and certain special objects. Conforming readers should read a PDF file from its end. The last line of the file shall contain only the end-of-file marker, **%%EOF**. The two preceding lines shall contain, one per line and in order, the keyword **startxref** and the byte offset in the decoded stream from the beginning of the file to the beginning of the **xref** keyword in the last cross-reference section. The **startxref** line shall be preceded by the *trailer dictionary*, consisting of the keyword **trailer** followed by a series of key-value pairs enclosed in double angle brackets (<<…>>) (using LESS-THAN SIGNs (3Ch) and GREATER-THAN SIGNs (3Eh)). Thus, the trailer has the following overall structure:

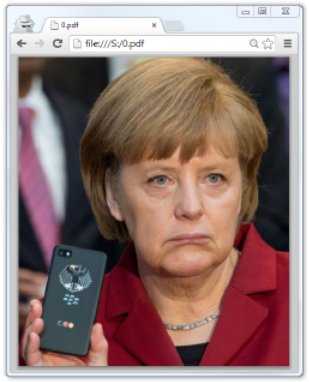
trailer
<< key_1 $value_1$
key_2 $value_2$
…
key_n $value_n$
>>
startxref
Byte_offset_of_last_cross-reference_section
%%EOF

```
% trailer <</Root …>>

trailer <</Root …>>

<</Root …>>
```

% trailer <</Root ...>>

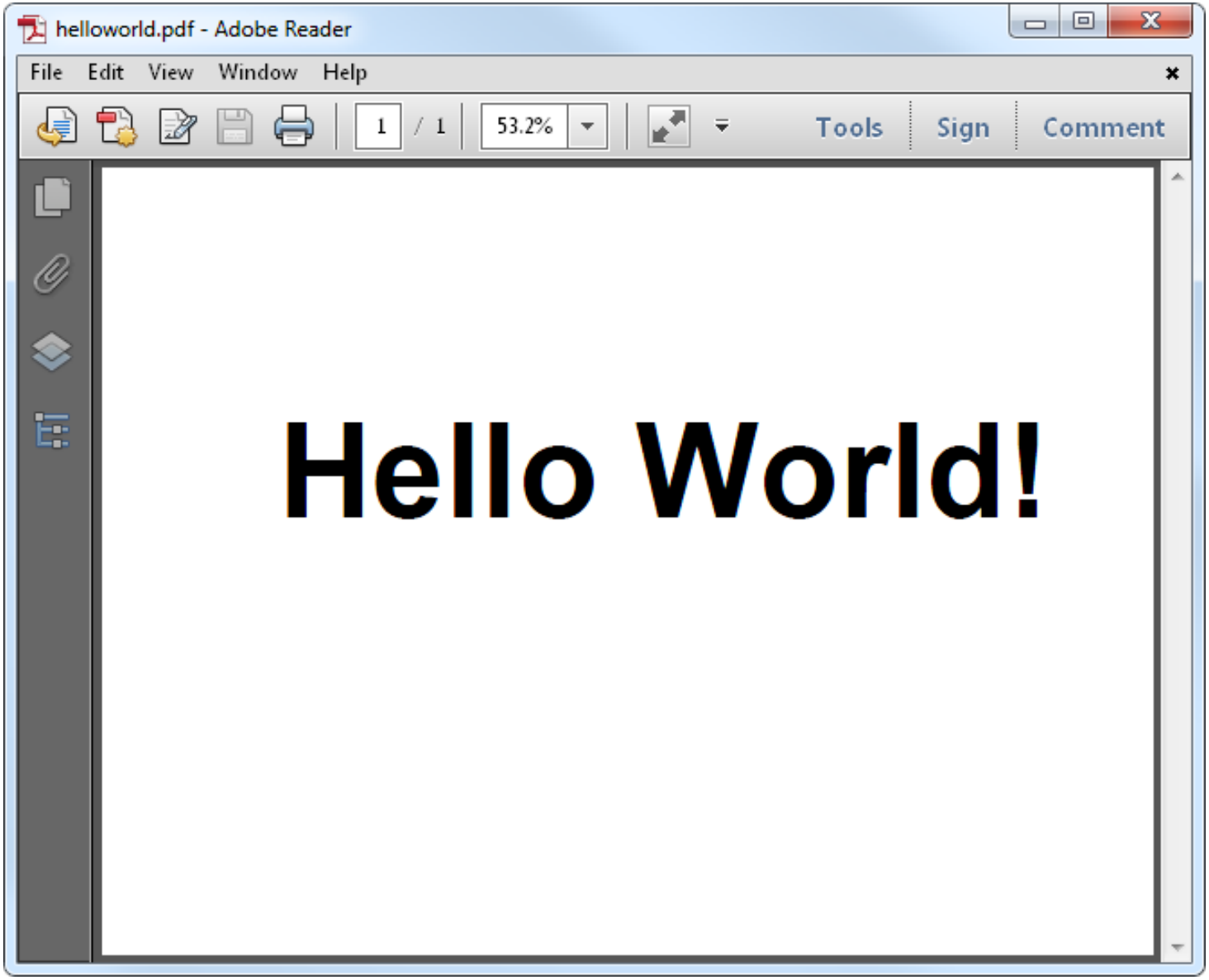
trailer <</Root ...>>

<</Root ...>>

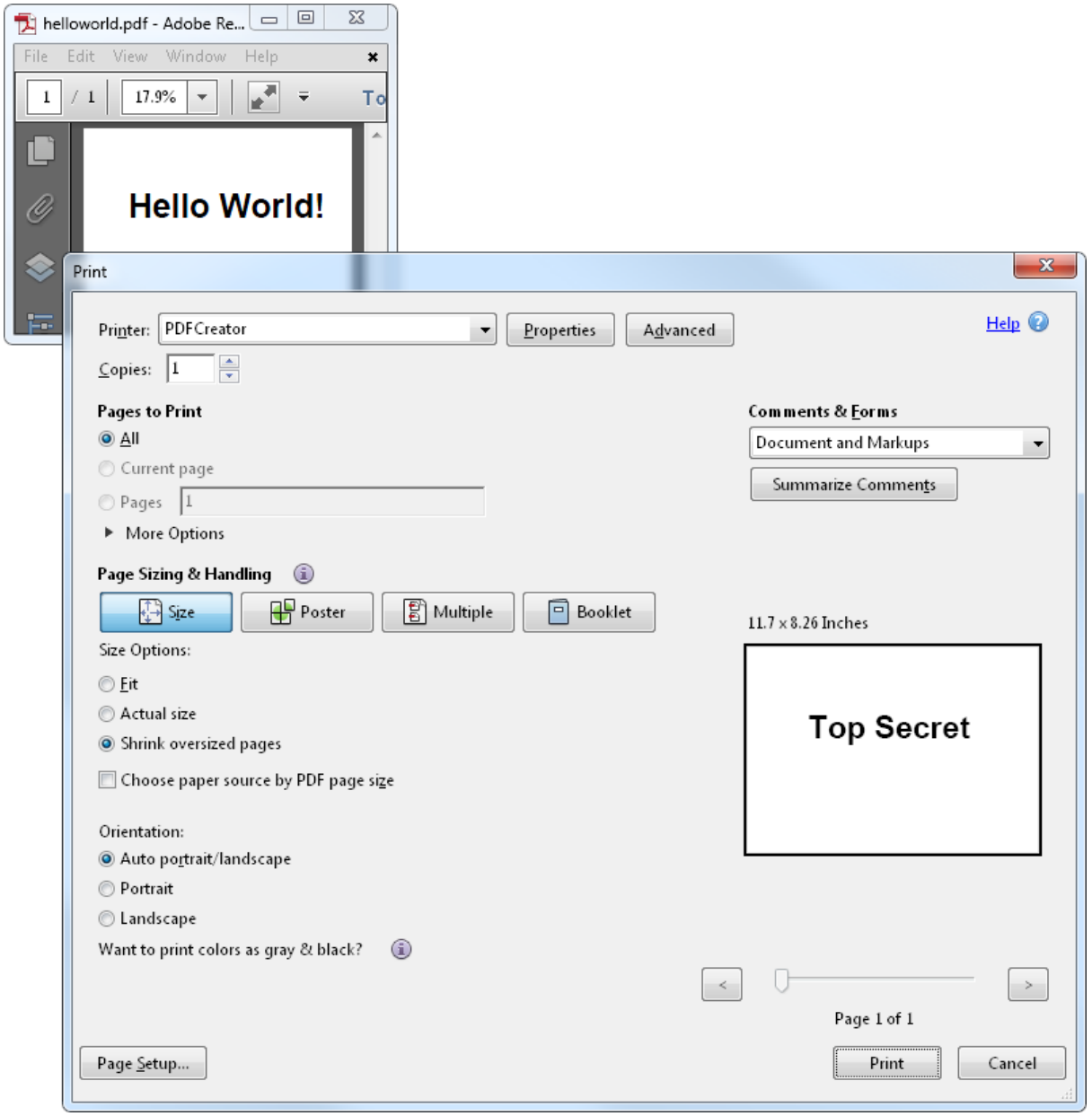


sometimes,
it’s in the specs

obscurity via over-specification?

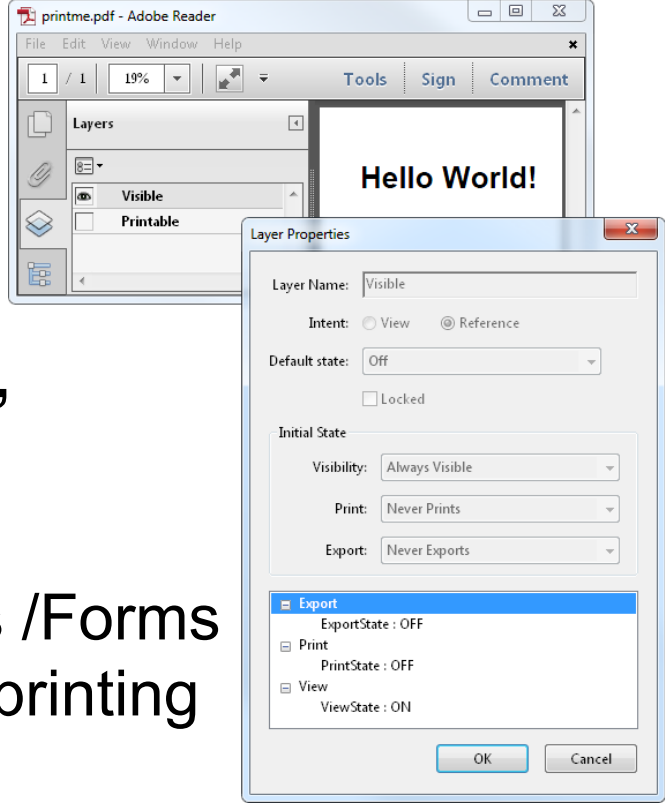


notice anything unusual?


helloworld.pdf - Adobe Re...
File Edit View Window Help
1 / 1
17.9%
Hello World!
Print
Printer: PDFCreator
Properties
Advanced
Help
Copies: 1
Pages to Print
All
Current page
Pages 1
More Options
Comments & Forms
Document and Markups
Summarize Comments
Page Sizing & Handling
Size
Poster
Multiple
Booklet
11.7 x 8.26 Inches
Size Options:
Fit
Actual size
Shrink oversized pages
Choose paper source by PDF page size
Top Secret
Orientation:
Auto portrait/landscape
Portrait
Landscape
Want to print colors as gray & black?
Page 1 of 1
Page Setup...
Print
Cancel

~~WYSIWYG~~

PDF Layers 1/2



“Optional Content Configuration”

- principles
  - define layered content via various /Forms
  - enable/disable layers on viewing/printing
- no warning when printing

- “you can see the preview!”
  - bypass preview by keeping page 1 unchanged
  - just do a minor change in the file

PDF Layers 2/2

- it’s Adobe only
  - what’s displayed varies with readers
  - could be hidden via previous schizophrenic trick
- it was in the specs all along
  - very rarely used
  - can be abused

BMP

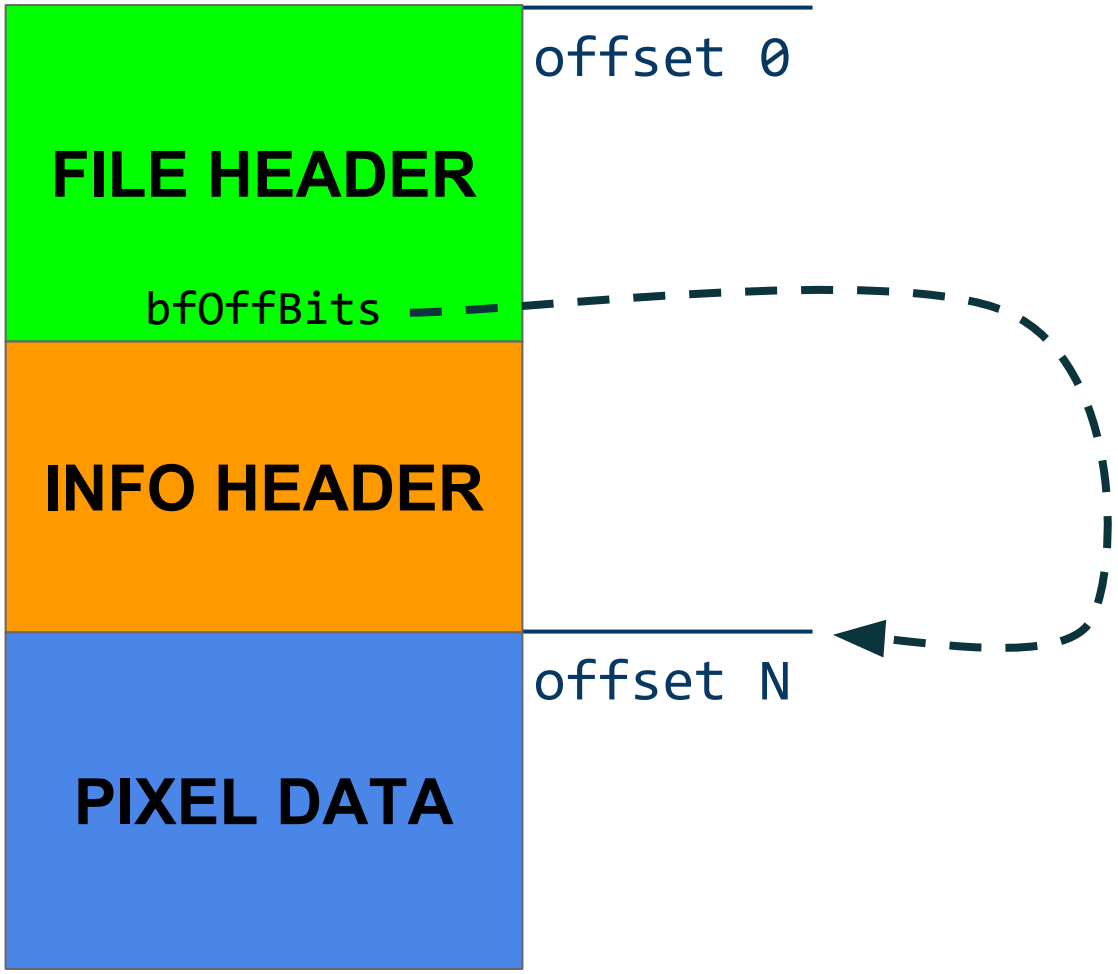
Trick 1

(originally published in Gynvael's "Format BMP okiem hakera" article in 2008)

bfOffBits

Specifies the offset, in bytes, from the BITMAPFILEHEADER structure to the bitmap bits

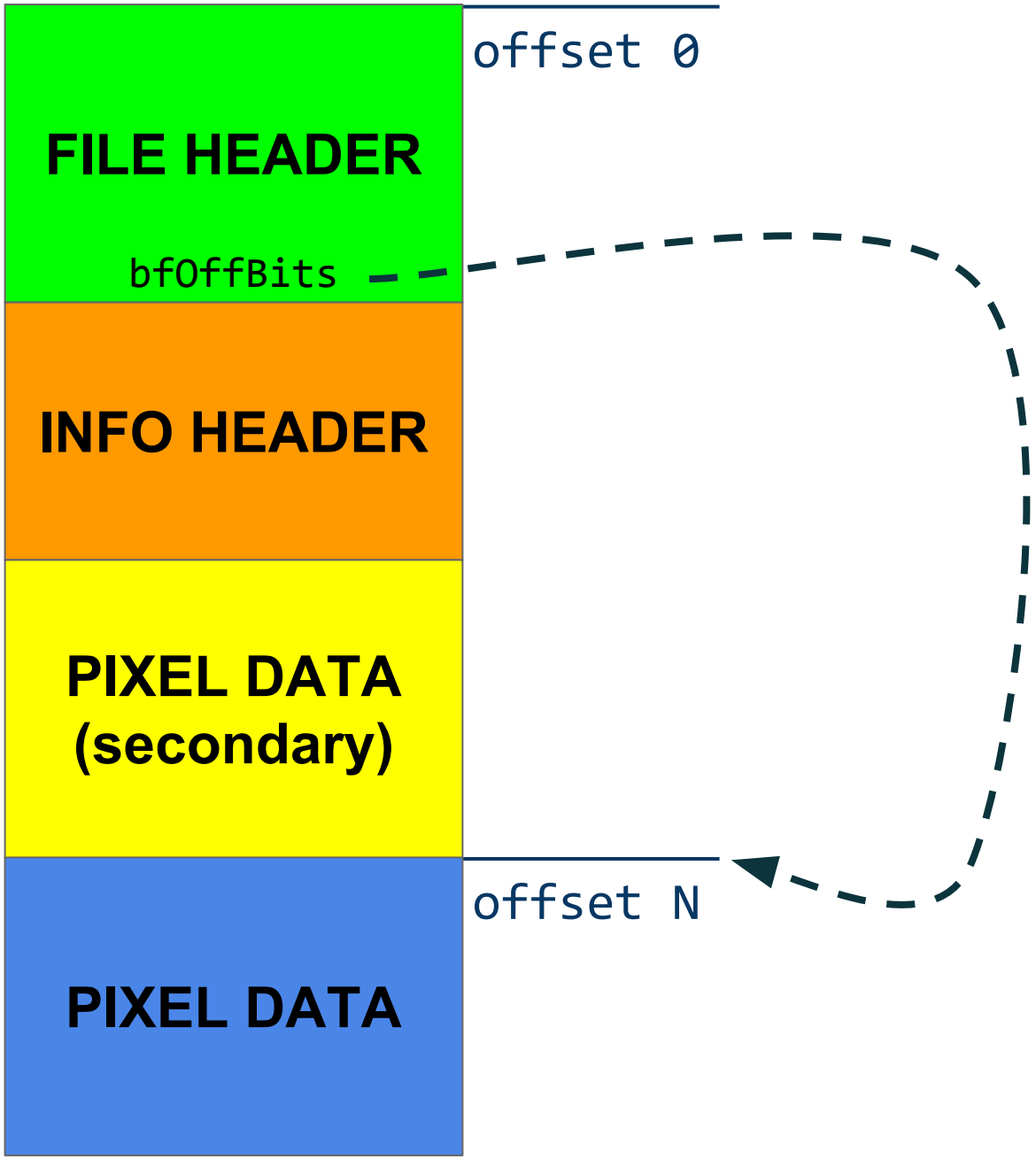
(MSDN)

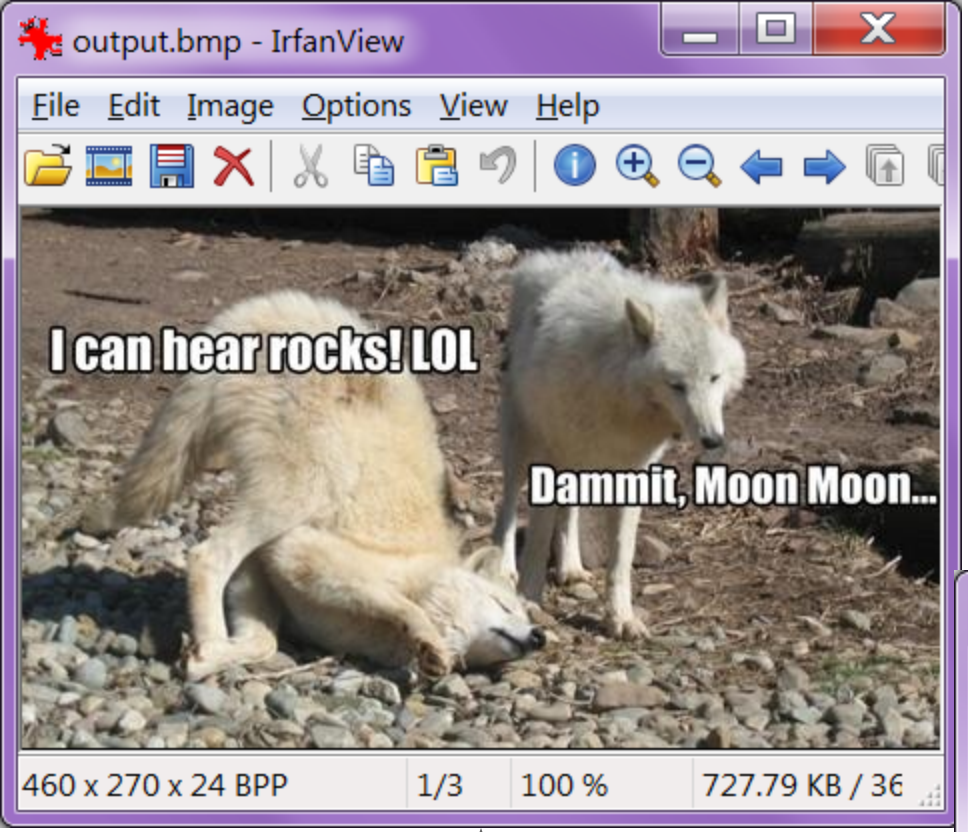


bfOffBits
Specifies the offset, in bytes, from the BITMAPFILEHEADER structure to the bitmap bits

(MSDN)

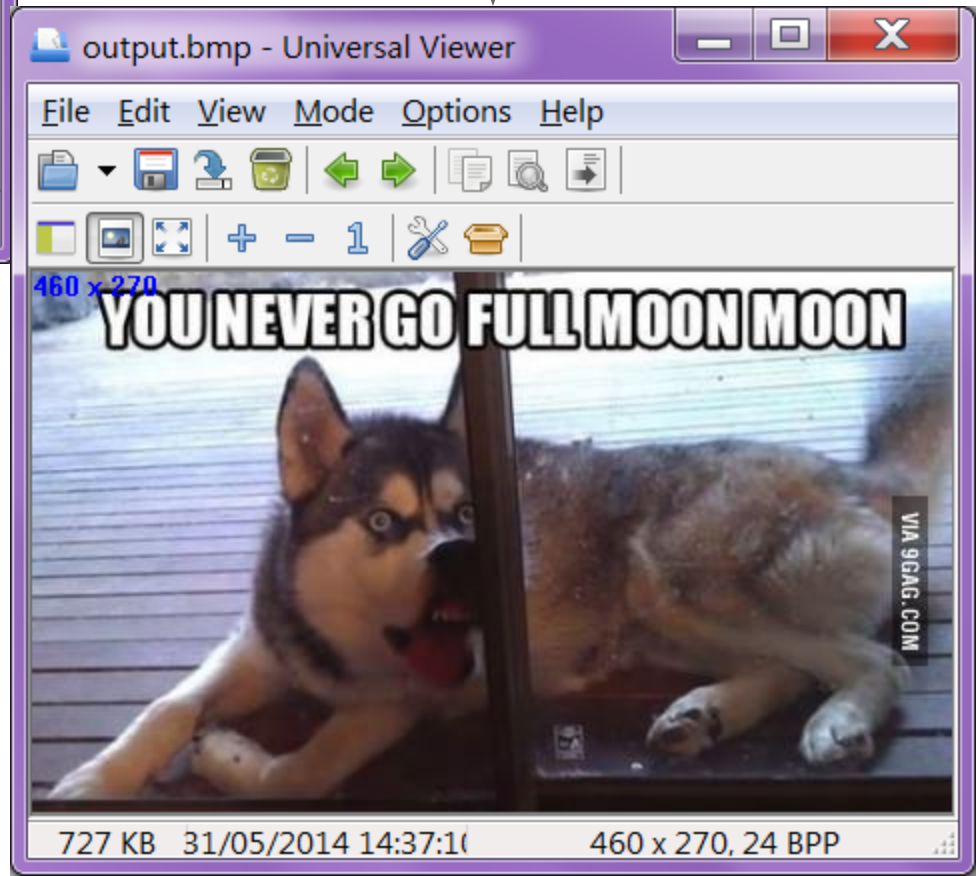
- Some image viewers ignore **bfOffBits** and look for data immediately after the headers.





PIXEL DATA

PIXEL DATA (secondary)



Different images, depending on which pixel data is used.

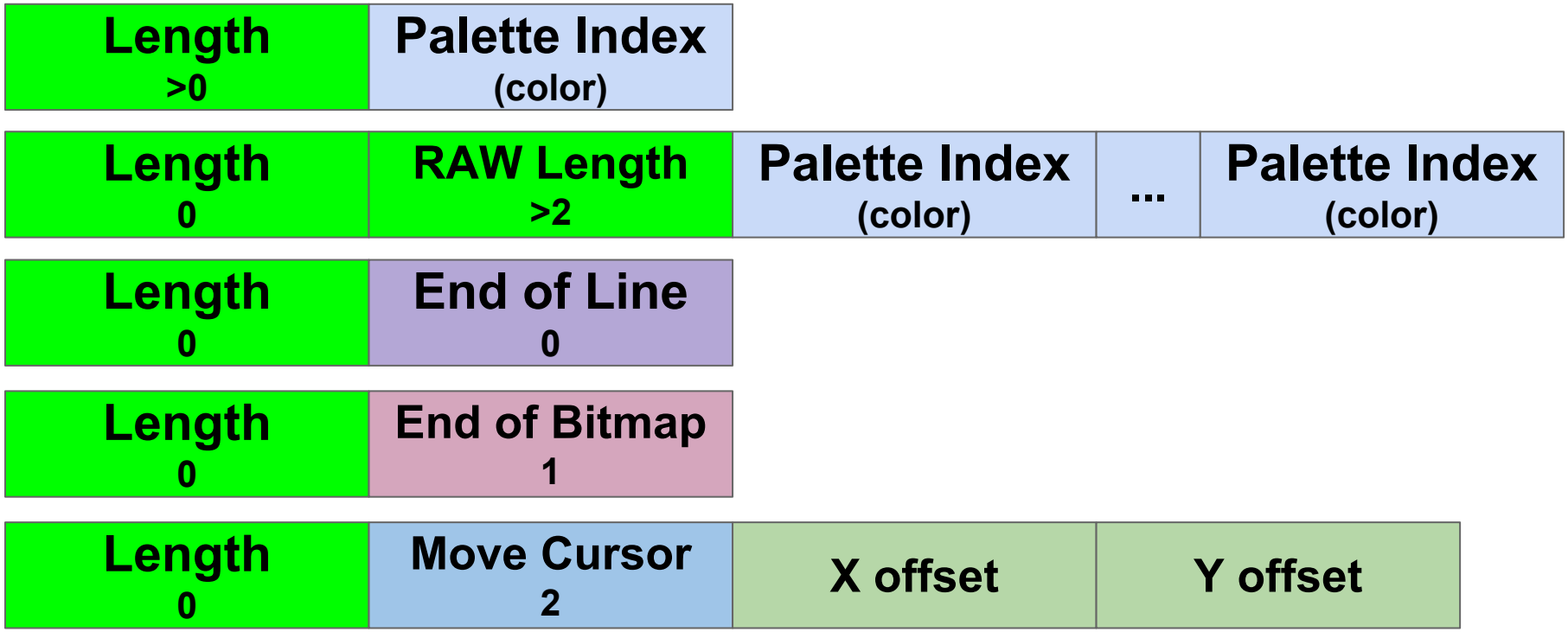
BMP

Trick 2

Something I've learnt about because it spoiled my steg100 task for a CTF (thankfully during testing).

BMP compression & palette

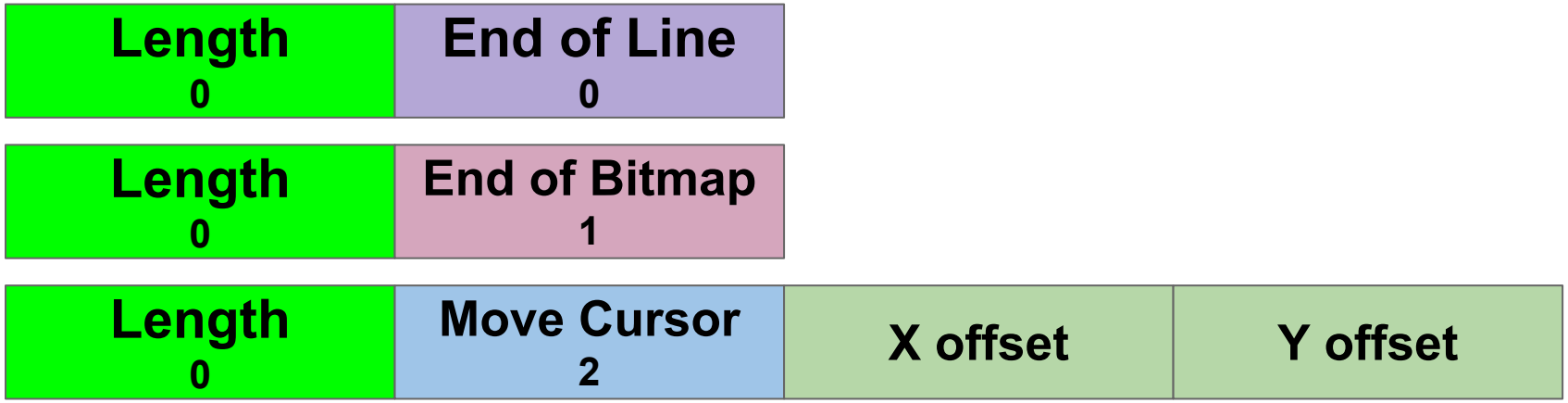
Run-Length Encoding (each box is **1** byte):



BMP compression & palette

Question: If the opcodes below allow jump over pixels and set no data, how will the pixels look like?

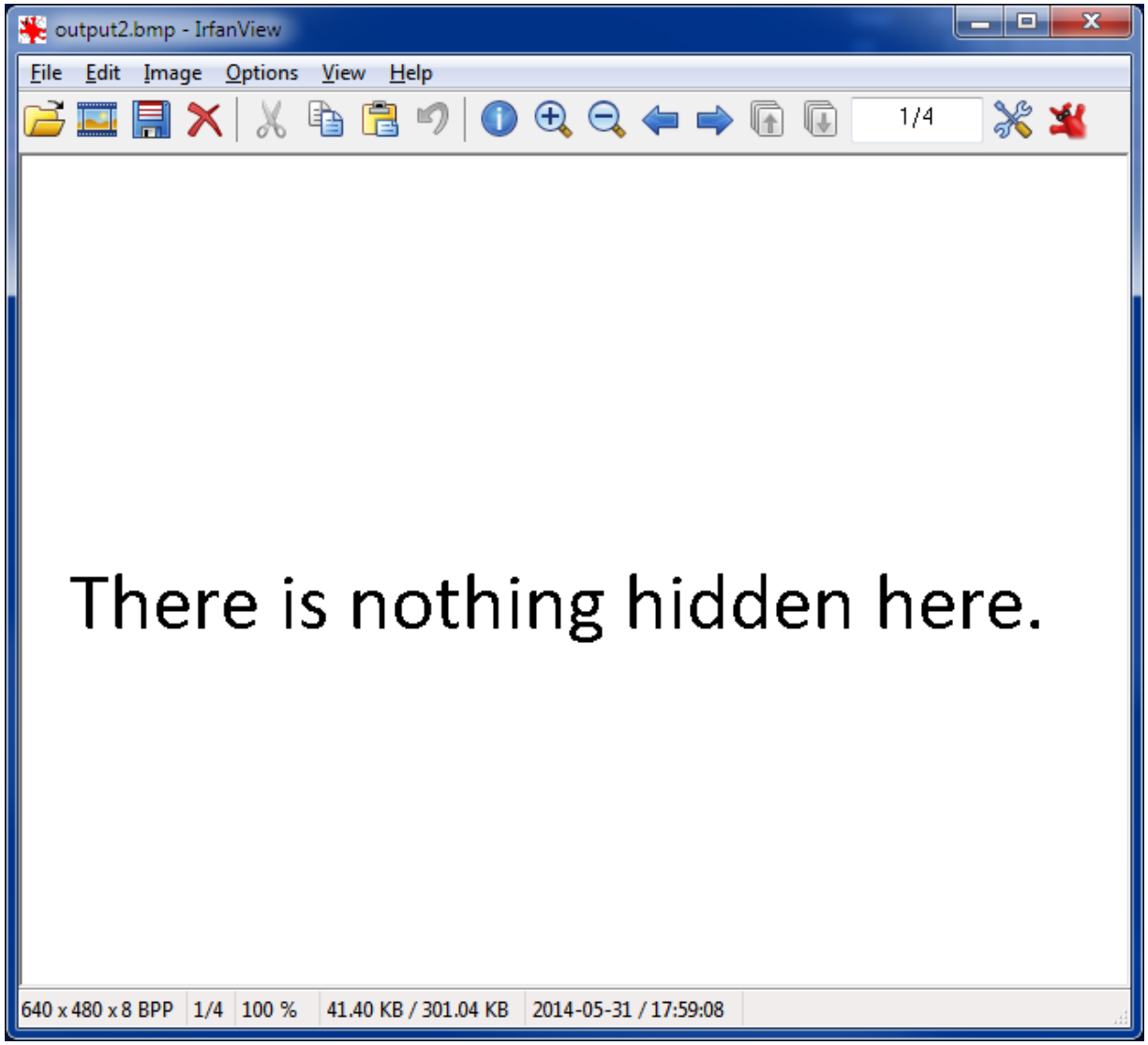
Hint: Please take a look at the presentation title :)



Option 1

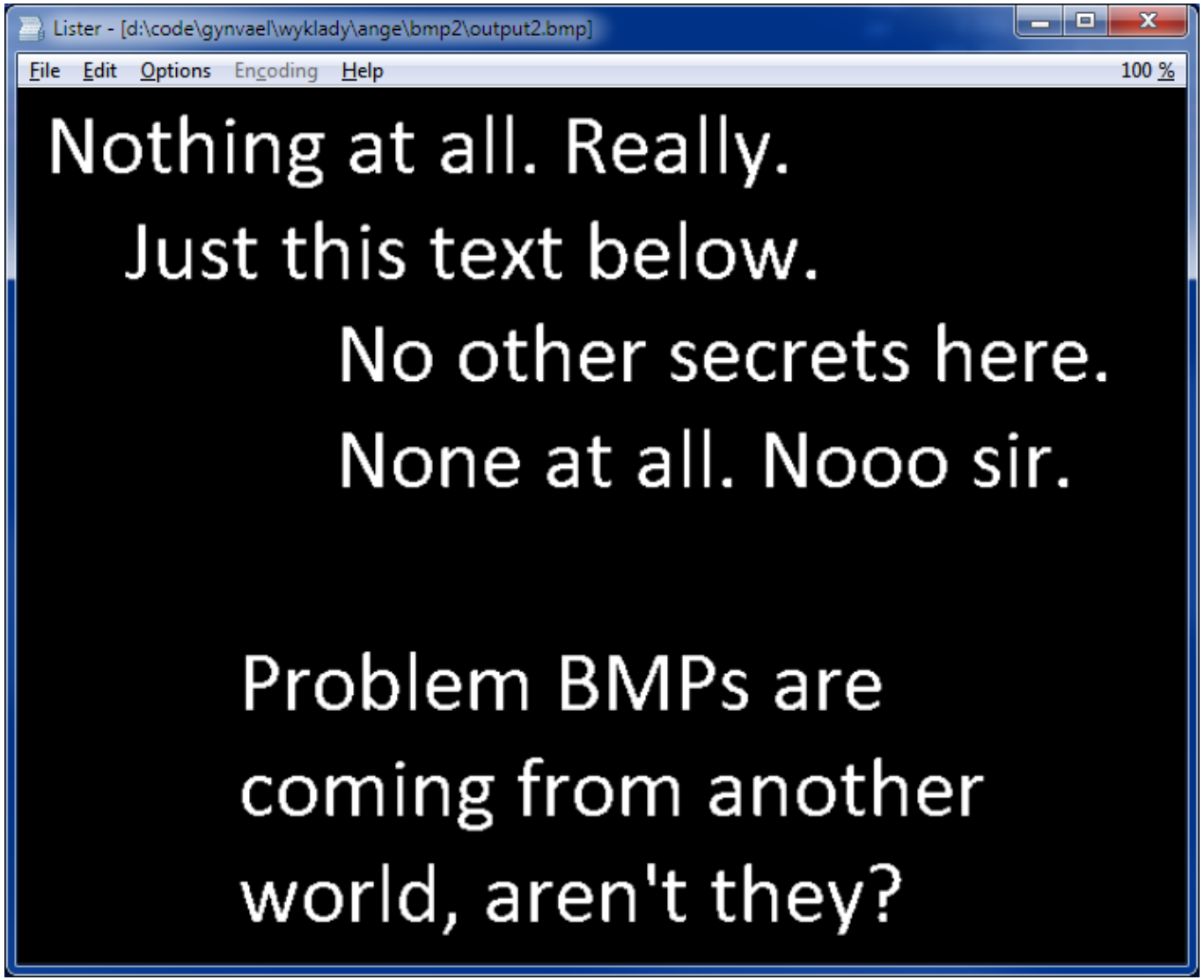
The missing data will be filled with **background color**.

(index 0 in the palette)



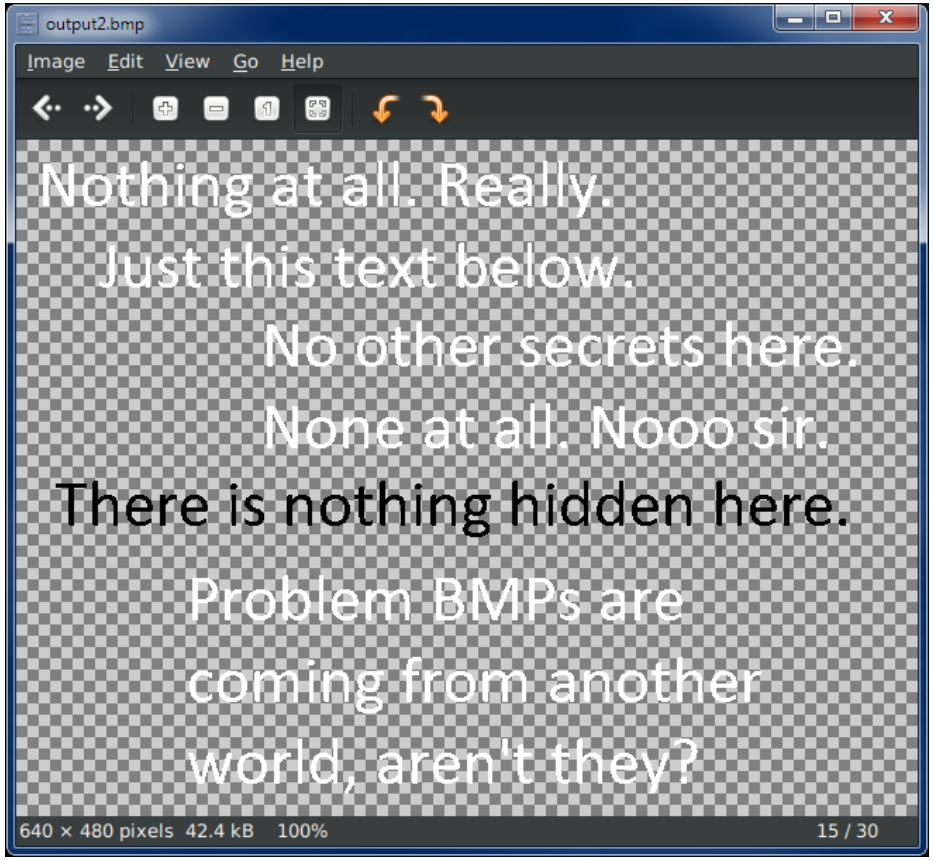
Option 2

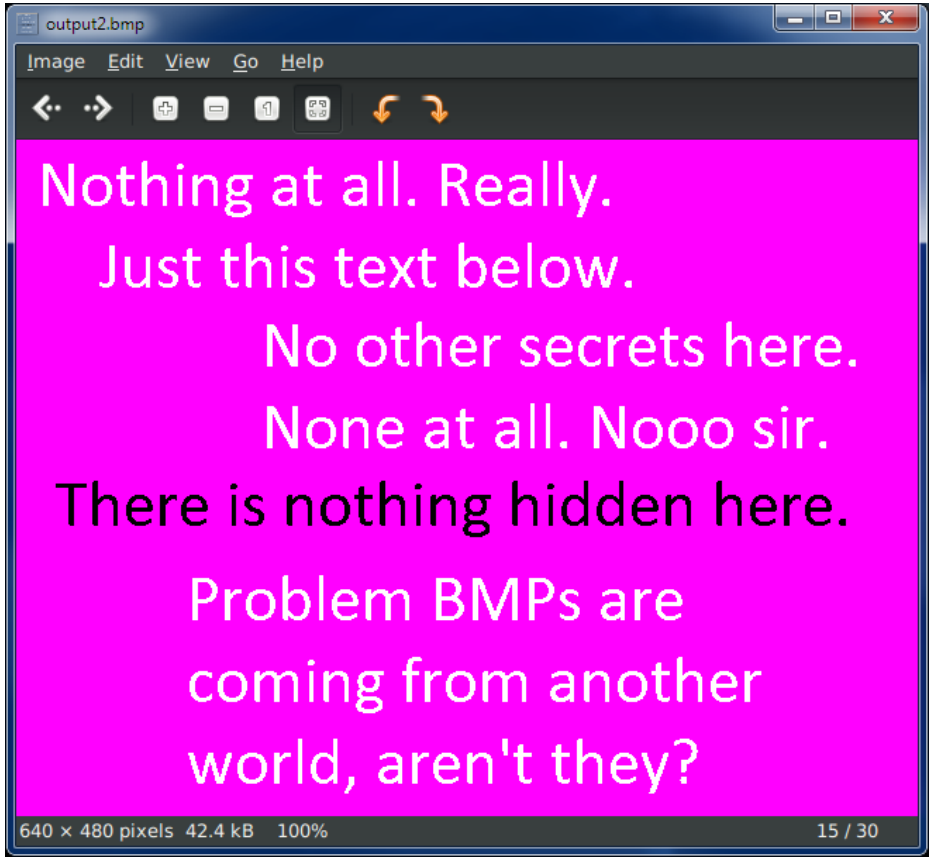
The missing data will be **black**.



Option 3

The missing data will be **transparent**.
(pink represents transparency)





PNG

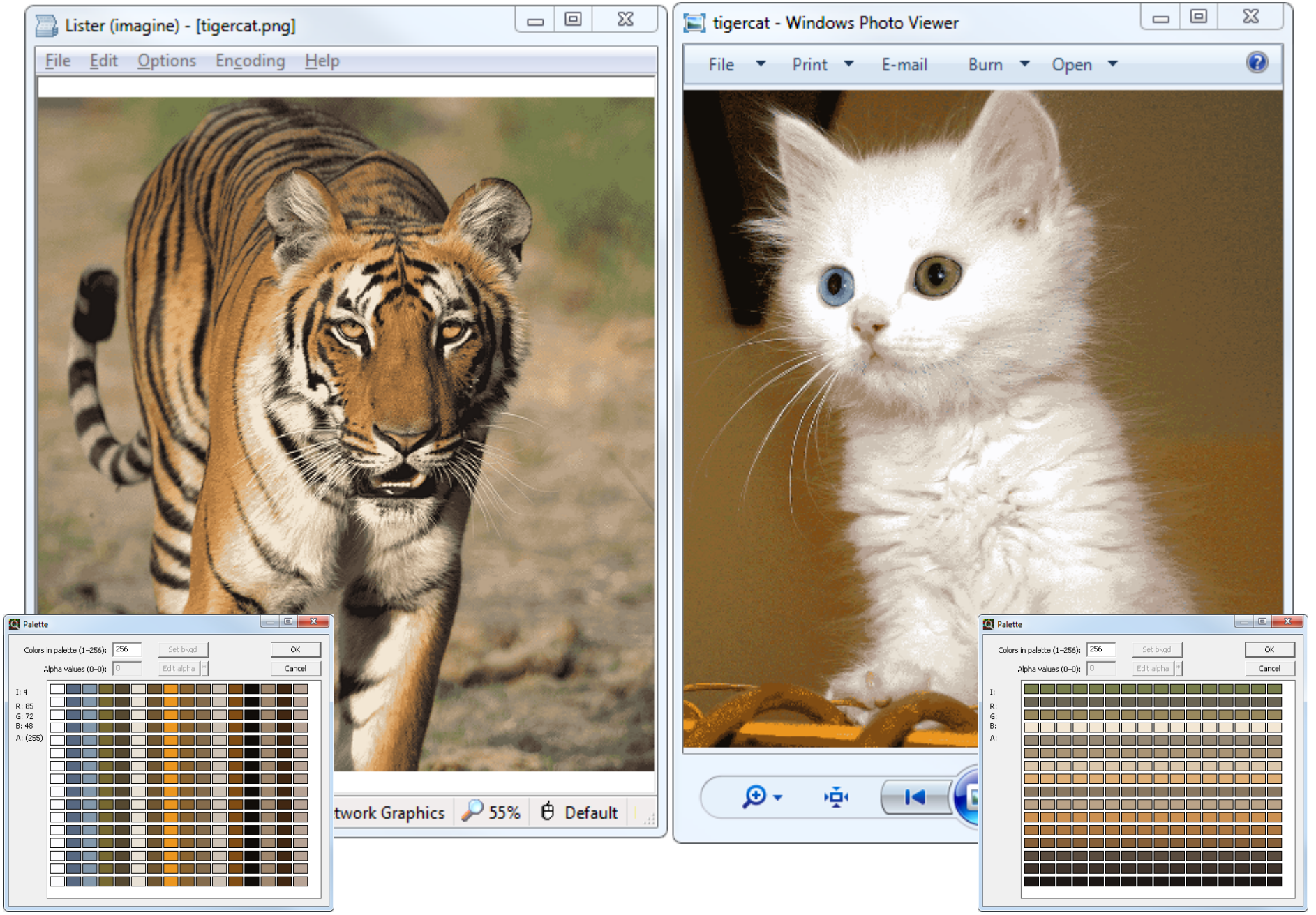
a data schizophren

image data combining

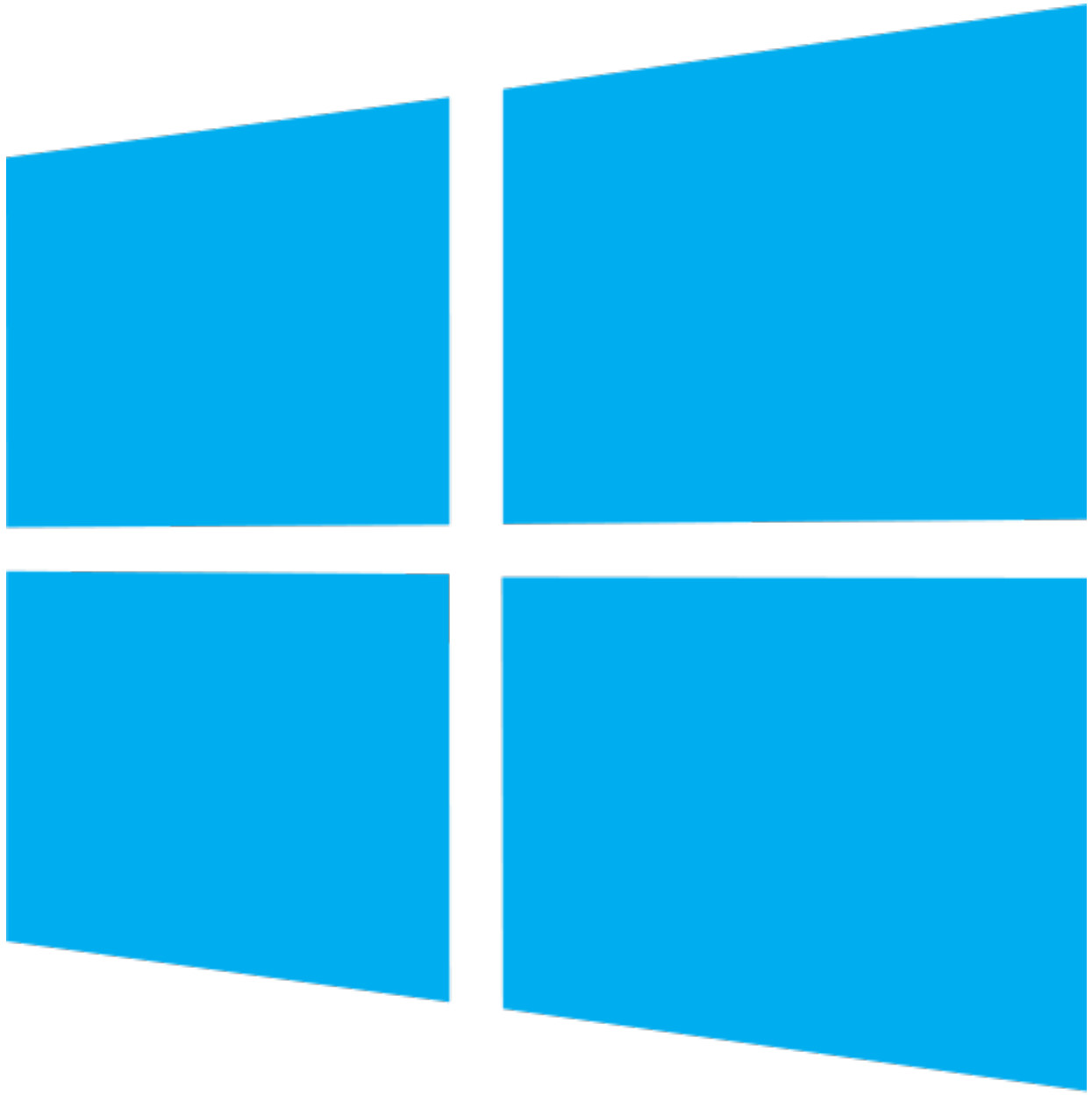
- 2 images
- via 2 palettes

cute PoC by **@reversity**

“There shall not be more than one PLTE chunk”



different images depending on which PLTE chunk is used



Relocations types



Type 4
HIGH_ADJ

--

--

Type 9
MIPS_JMPADDR16
IA64_IMM64
MACHINE_SPEC_9

32 bit

64 bit

as seen in PoC||GTFO

Relocations on relocations

Type 4
HIGH_ADJ

--

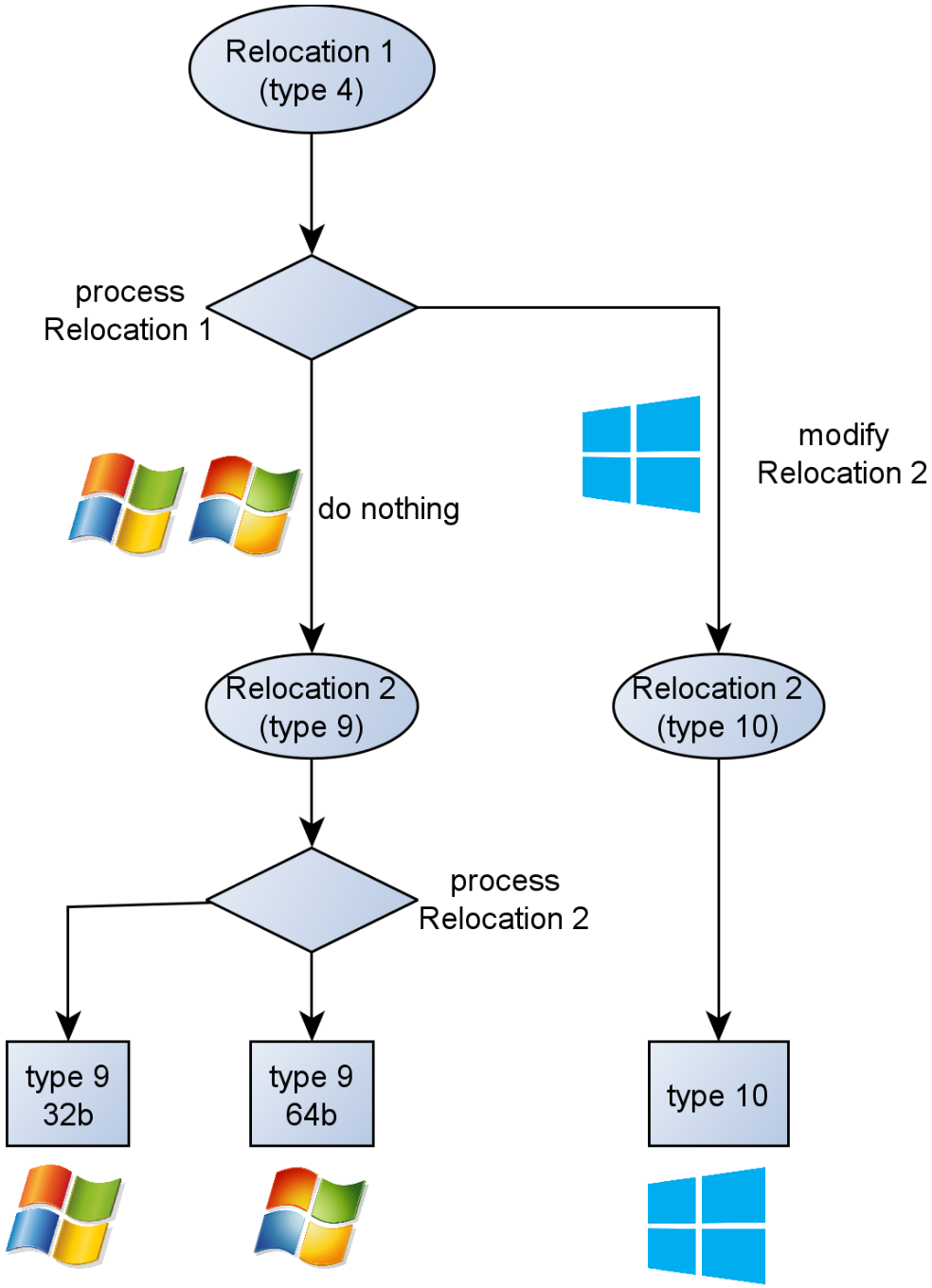
--

Type 9
MIPS_JMPADDR16
IA64_IMM64
MACHINE_SPEC_9

32 bit

64 bit

Type 10
DIR64



Relocation-based PE Schizophren

Julian Bangert, Sergey Bratus -- ELF Eccentricities
https://www.youtube.com/watch?v=4LU6N6THh2U

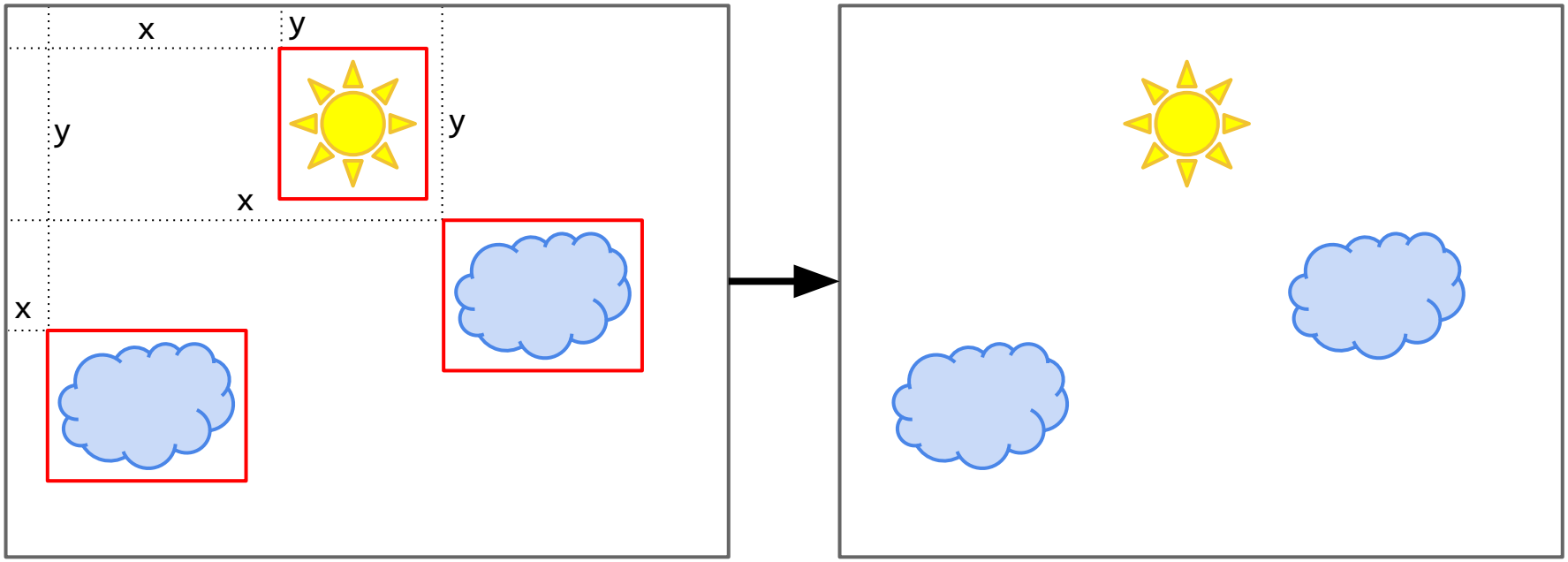
GIF

Something Gynvael stumbled on in 2008,
but never made a PoC... until now.
(with great input from Ange)

GIF

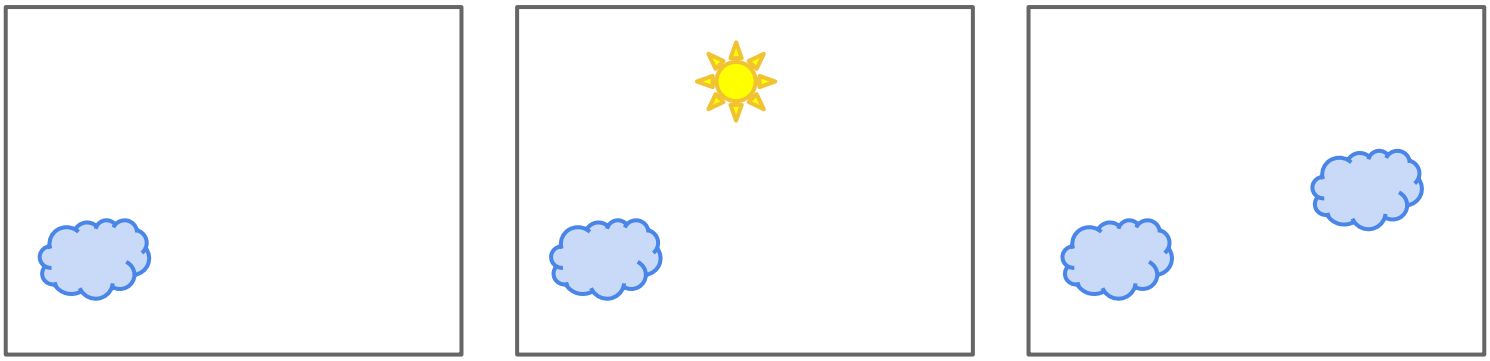
GIF can be made of many small images.

If "frame speed" is defined, these are frames instead (and the first frame is treated as background).



GIF

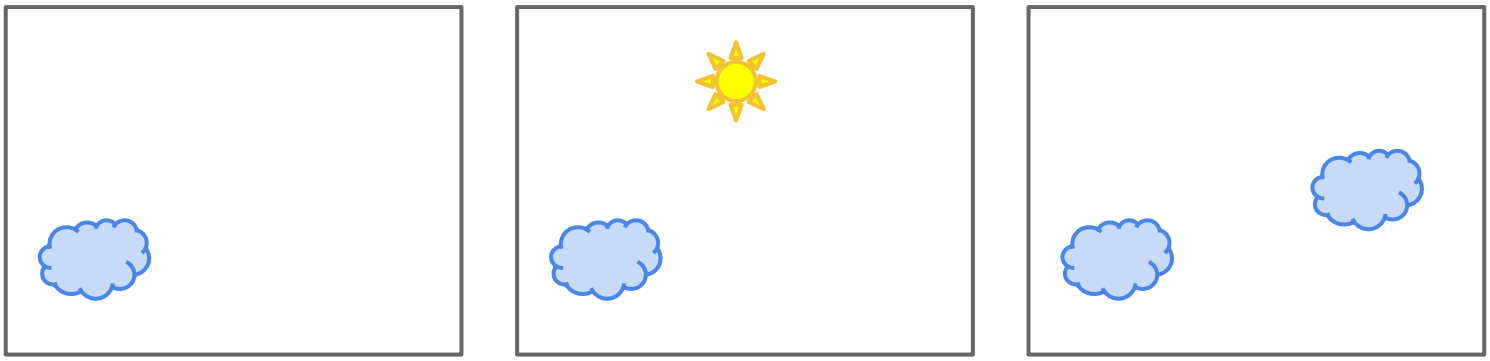
Certain parsers (e.g. browsers) treat "images" as "frames" regardless of "frame speed" not being defined.



Frame 1 Frame 2 Frame 3

GIF

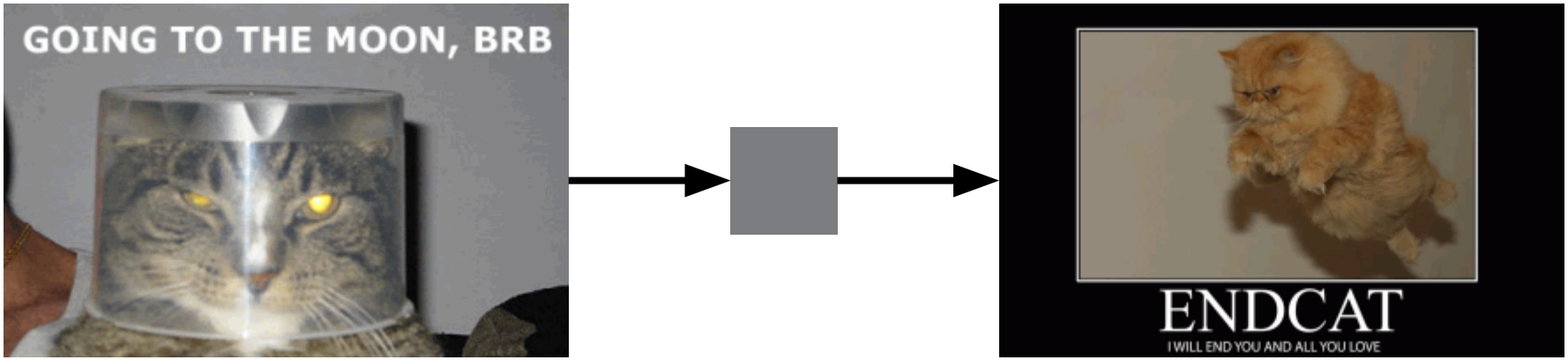
Certain parsers (e.g. browsers) treat "images" as "frames" regardless of "frame speed" not being defined.



Frame 1 Frame 2 Frame 3

GIF

Schizophrenic PoC:



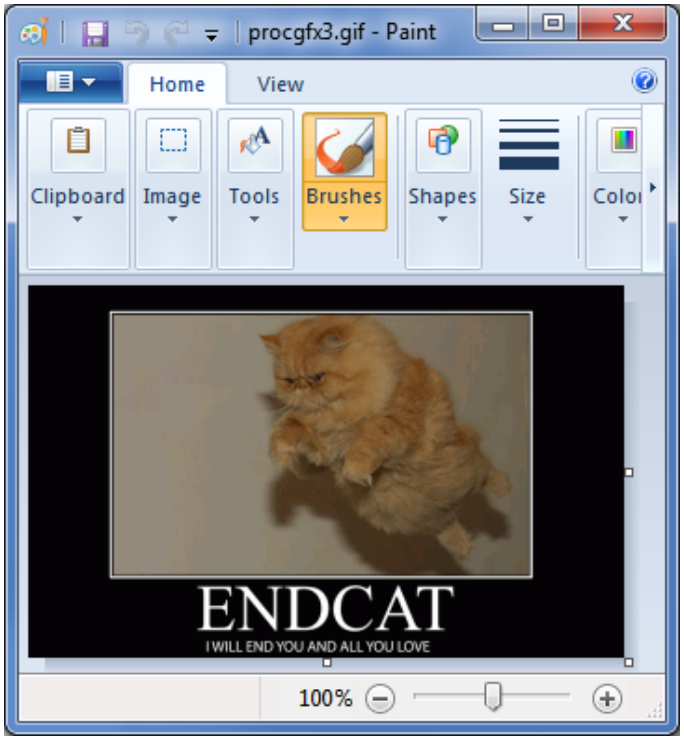
Frame 1

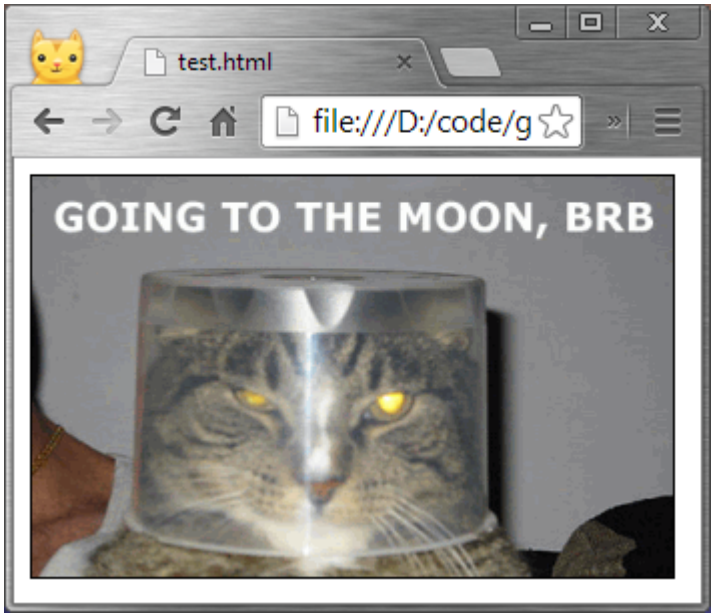
Frames 2-10001

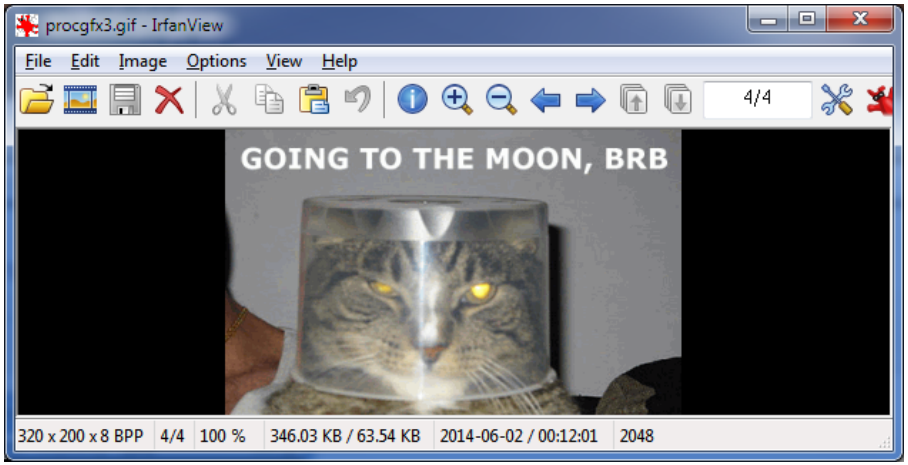
1x1 px

Frame 10002

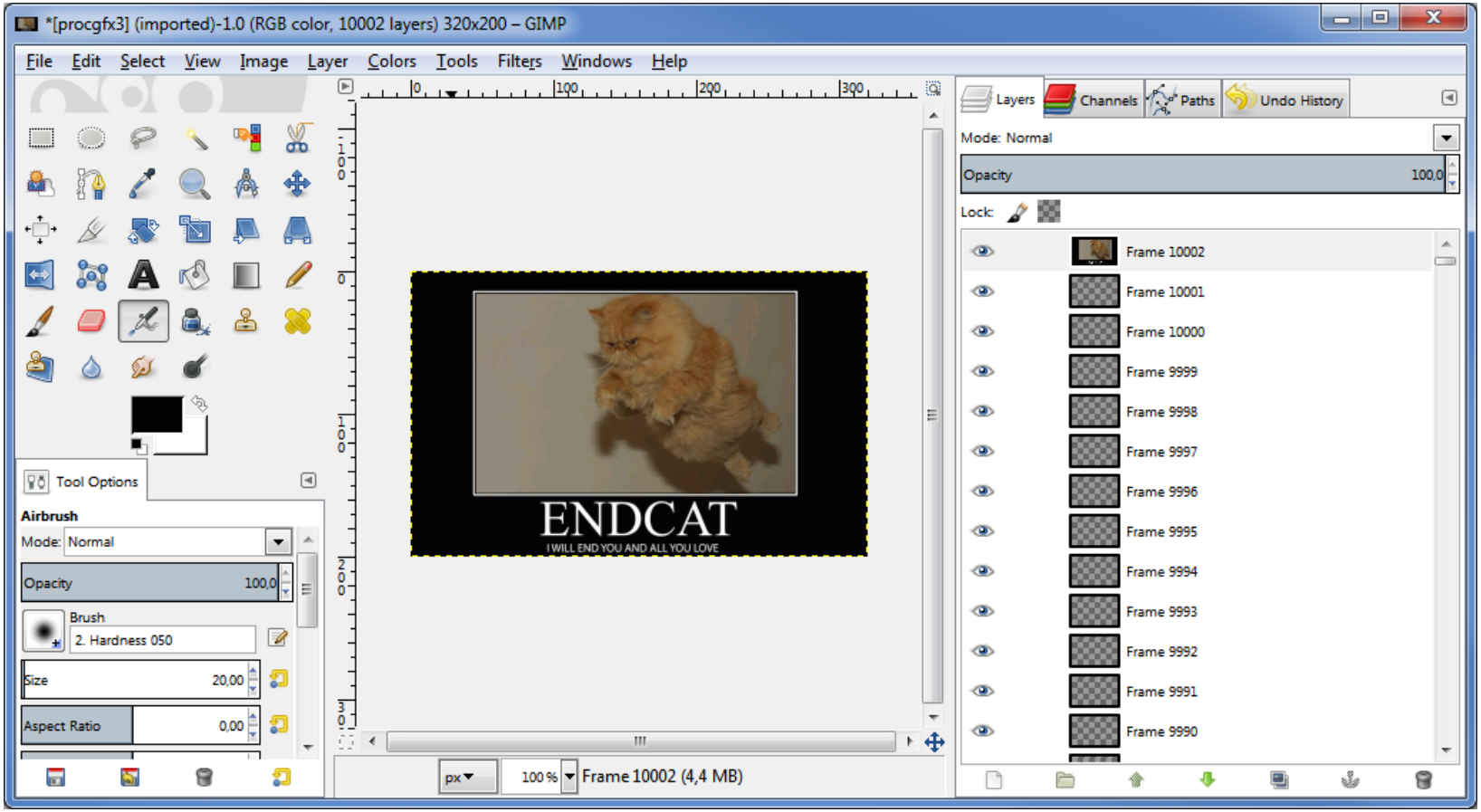
These apps render the GIF by the specs.



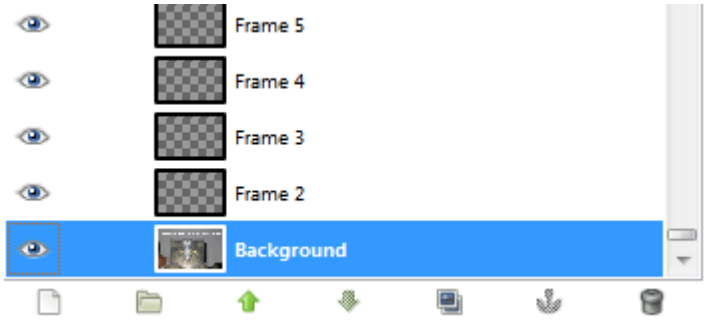




These apps try to force animation.



GIMP says "frames", but allows one to see all the frames, which is nice.



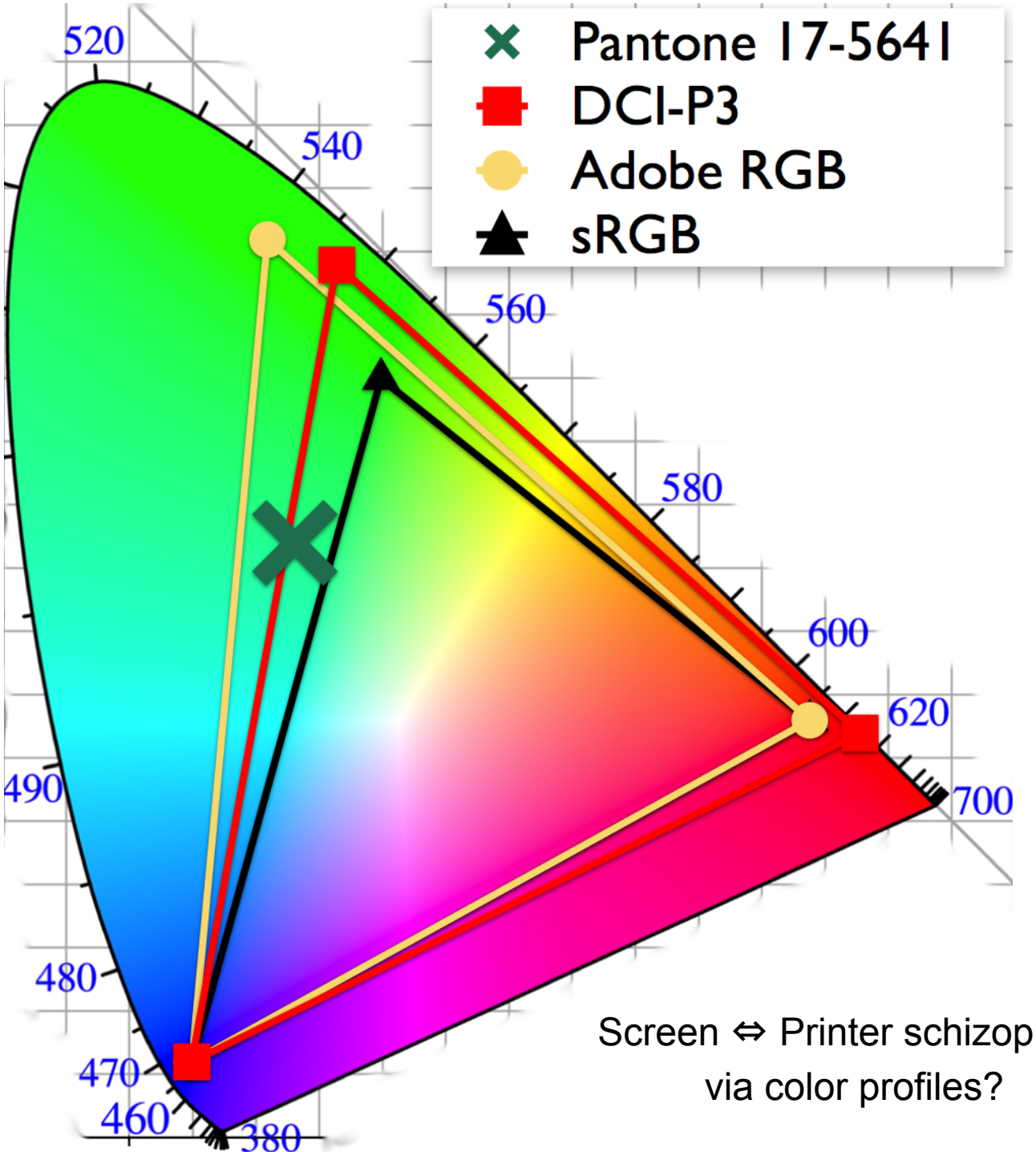
same-tool schizophrenia

1 file + 1 tool = 2 behaviors

it was too simple

- WinRar: different behavior when viewing or extracting
  - opening/failing
  - opening/’nothing’
- Adobe: viewing ⇔printing
  - well, it’s a *feature*

Failures / Ideas / WIP



Screen ⇔ Printer schizophren
via color profiles?

Failures / Ideas / WIP

- screen ⇔ printer
  - embedded color profiles?
- JPG
  - IrfanView vs the world
- Video
  - FLV: video fails but still plays sound ?

PNG

Various ancillary chunks (rendering level)

- partially supported:
  - gamma
  - transparency (for palettes)
- never supported?
  - significant bits
  - chromacities
- always supported?
  - physical size

Conclusion

Conclusion

- such a mess
  - specs are messy
  - parsers don’t even respect them
- no CVE/blaming for parsing errors?
  - no security bug if no crash or exploit :(

PoCs and slides: http://goo.gl/Sfjfo4

ACK

@reversity @travisgoodspeed @sergeybratus
qkumba @internot @pdfkungfoo

@j00ru ise ds vx

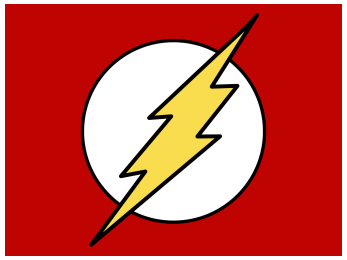
thank you

@angealbertini
@gynvael

AREA41

Bonus Round

(not a fully schizophrenic problem in popular parsers, that's why it's here)



vs

Prezi SWF sanitizer

Prezi allows embedding SWF files.

But it first sanitizes them.

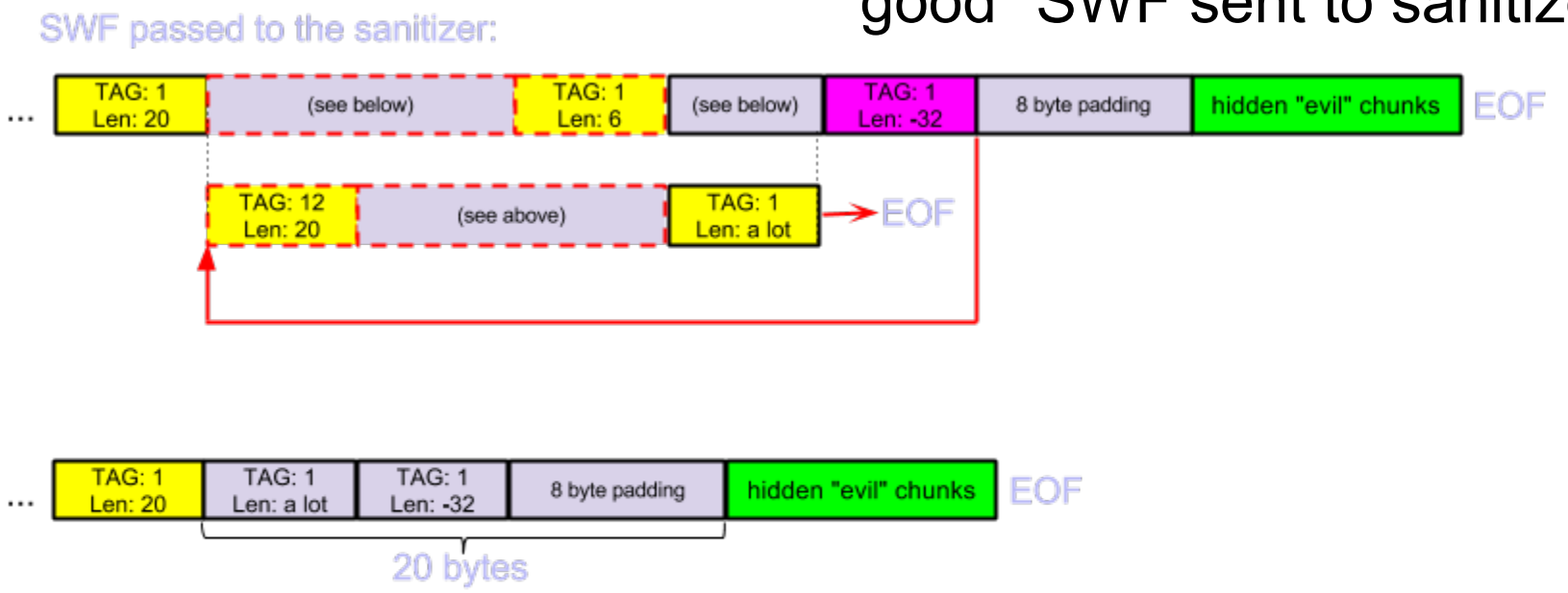
It uses one of two built-in SWF parsers.

There **was** a problem in one of them:

- It allowed huge chunk sizes.
- It just "jumped" (seeked) over these chunk...
- ...which resulted in an integer overflow...
- ...and this lead to schizophrenia.
- As the sanitizer saw a good SWF...
- ...Adobe Flash got its evil twin brother.

Prezi SWF sanitizer

"good" SWF sent to sanitizer



and its evil twin brother
kudos to the sanitizer!

Fixed in Q1 2014. For details see:
"*Integer overflow into XSS and other fun stuff - a case study of a bug bounty*"
http://gynvael.coldwind.pl/?id=533