

Hachoir

A person wearing a blue long-sleeved shirt is using a large, industrial-style meat chopper. The person's hands are on the handles of the device, which is a tall, rectangular metal box with a circular vent on the front. The background is a dimly lit industrial or kitchen setting with concrete walls and a fluorescent light fixture.

« Attention, ça va trancher chérie »

AAM Janvier 2007 - Victor Stinner

Hachoir core

- ◆ Tailles et adresses en bit
- ◆ Chaînes de caractères Unicode
- ◆ S'occupe de l'ordre des octets (*endian*)
- ◆ Nombreux types prédéfinis
- ◆ Orienté objet (ex: un champ est un objet)
- ◆ Parseur paresseux
- ◆ Correction automatique en cas d'erreur


Hachoir parser

- ♦ **Image** : BMP, CUR, EMF, GIF, ICO, JPEG, PCX, PNG, PSD, TGA, TIFF, XCF, WMF
- ♦ **Audio** : AIFF, AIFC, iTunes, MIDI, MPEG, OGG/Vorbis, Real, Sun/NeXT, WMA
- ♦ **Vidéo** : FLV, Matroska, MPEG, MOV, Ogg/Theora, Real, WMV
- ♦ **Archive** : 7-zip, bzip2, Debian, gzip, RPM, TAR, ZIP
- ♦ **Système de fichier** : EXT2, EXT3, FAT12, FAT16, FAT32, Linux swap, ISO 9660, MBR, ReiserFS3
- ♦ **Metadonnées** : 8BIM, AMF, EXIF, ID3, IPTC
- ♦ **Programme** : CLASS, ELF, EXE, PYC, PYO
- ♦ **Divers** : 3DO, 3DS, ASN.1, MS Office, SWF, Tcpdump

Hachoir parser

- ◆ Parseurs plus ou moins complets
- ◆ Décompression à la volée
- ◆ Choix automatique du parseur
- ◆ Possibilité de parser un fichier contenu dans un autre fichier

Hachoir urwid



0) file:/home/haypo/testcase/KDE_Click.wav: Microft WAVE audio (1824 bytes)
0) signature= "RIFF": AVI header (RIFF) (4 bytes)
4) filesize= 1816 bytes: File size (4 bytes)
8) type= "WAVE": Content type ("AVI ", "WAVE", ...) (4 bytes)
- 12) format: Audio format (24 bytes)
0) tag= "fmt ": Tag (4 bytes)
4) size= 16 bytes: Size (4 bytes)
8) codec= Microsoft Pulse Code Modulation (PCM): Audio codec (2 bytes)
10) nb_channel= 2: Number of audio channel (2 bytes)
12) sample_per_sec= 22050: Sample per second (4 bytes)
16) byte_per_sec= 88200: Average byte per second (4 bytes)
20) block_align= 4: Block align (2 bytes)
22) bit_per_sample= 16: Bits per sample (2 bytes)
+ 36) audio_data: Audio stream data (1732 bytes)
- 1768) info: File informations (56 bytes)
0) tag= "LIST": Tag (4 bytes)
4) size= 48 bytes: Size (4 bytes)
8) subtag= "INFO": Sub-tag (4 bytes)
- 12) creation_date: Creation date (20 bytes)
0) tag= "ICRD": Tag (4 bytes)
4) size= 11 bytes: Size (4 bytes)
8) text= "2001-02-21" (11 bytes)
19) padding[0]= "\0" (1 byte)
- 32) producer: Producer (24 bytes)
0) tag= "ISFT": Tag (4 bytes)
4) size= 16 bytes: Size (4 bytes)
8) text= "Sound Forge 4.5" (16 bytes)

Hachoir wx

/home/haypo/testcase/logo-Kubuntu.png/header

File

```
89 50 4e 47 0d 0a 1a 0a 00 00 00 0d 49 48 44 52 00 00
01 4b 00 00 00 5a 08 06 00 00 00 31 4a 74 14 00 00 00
06 62 4b 47 44 00 ff 00 ff 00 ff a0 bd a7 93 00 00 00
09 70 48 59 73 00 00 0b 13 00 00 0b 13 01 00 9a 9c 18
00 00 00 07 74 49 4d 45 07 d6 05 1a 09 29 2e 68 b6 c4
```

address	name	type	size	data
	../			
00000000.0	size	UInt32	00000004.0	13
00000004.0	tag	FixedString	00000004.0	"IHDR"
00000008.0	width	UInt32	00000004.0	331
00000012.0	height	UInt32	00000004.0	90
00000016.0	bpp	UInt8	00000001.0	8
00000017.0	reserved	RawBits	00000000.5	0
00000017.5	alpha	Bit	00000000.1	True
00000017.6	color	Bit	00000000.1	True
00000017.7	palette	Bit	00000000.1	False
00000018.0	compression	UInt8	00000001.0	deflate
00000019.0	filter	UInt8	00000001.0	0
00000020.0	interlace	UInt8	00000001.0	0
00000021.0	crc32	UInt32	00000004.0	0x314a7414

Hachoir metadata



```
$ hachoir-metadata matrix_ping_pong.wmv
Common:
- Title: 欽ちゃん&香取慎吾の全日本仮装大賞
- Author: Nippon Television Network Corporation[NTV]
- Duration: 1 min 47 sec
- Creation date: 2003-06-16 07:57:23.235000
- Copyright: [C]Nippon Television Network Corporation[NTV] 2003
- Comment: Is seekable
- Comment: Max bit rate: 276.9 Kbit/sec
- Comment: WMFSDKVersion=7.01.00.3055
- Comment: WMFSDKNeeded=0.0.0.0000
- MIME type: video/x-ms-wmv
- Endian: Little endian
Audio stream #1:
- Sample rate: 8.0 KHz
- Bits/sample: 16 bits
- Compression: Windows Media Audio V7 / V8 / V9
- Bit rate: 13.0 Kbit/sec
Video stream #1:
- Image width: 200
- Image height: 150
- Bits/pixel: 24
- Compression: Windows Media Video V7
- Bit rate: 16.3 Kbit/sec
Video stream #2:
- Image width: 200
- Image height: 150
- Bits/pixel: 24
- Compression: Windows Media Video V7
- Bit rate: 36.3 Kbit/sec
Video stream #3:
- Image width: 200
- Image height: 150
- Bits/pixel: 24
- Compression: Windows Media Video V7
- Bit rate: 211.3 Kbit/sec
```


Hachoir metadata

- ◆ Informations ordonnées par importance
- ◆ Données lisibles, ex: « 1 hour 26 min 52 sec » et non « 5160 sec »
- ◆ Identifiants communs, ex: duration, compression, producer, ...
- ◆ Possibilité de filtrer les informations

Hachoir subfile

- ◆ Recherche les fichiers contenus dans un autre fichier à n'importe quelle position
- ◆ Cherche le début d'un fichier (utilisation de motifs pour accélérer la recherche)
- ◆ Valide le fichier en vérifiant certaines valeurs importantes
- ◆ Tente de calculer la taille du fichier

Hachoir subfile

- ◆ Ne supporte pas la fragmentation
- ◆ Projets similaires : *PhotoRec* et *Scalpel*
(réécriture de *Foremost*)
- ◆ Détecte encore beaucoup de faux positifs

Idées

- ◆ Reconnaissance automatique de motifs pour deviner un format de fichier (ingénierie inverse)
- ◆ Intégration d'hachoir-metadata dans Nautilus, Konqueror, Beagle, etc.
- ◆ Fuzzer basé sur Hachoir
- ◆ hachoir-strip : suppression des méta-données



C'est fini !

Des questions ?