Forensic Analysis of Deduplicated File Systems

Dario Lanterna

Antonio Barili

Digital Forensics Lab
Dept. of Industrial and Information Engineering
University of Pavia (Italy)
labinfor@unipv.it

Why study the storage deduplication?

- Deduplication is a technology that reduces space used on storage devices. Backup is one of the most important field of application.
- Investigations starts after the crime is committed, therefore data from backups is an important source of evidences;
- There are many implementation of deduplication. I focus attention on **OpenDedup** and **Microsoft** implementations.

Where can I find Deduplication?

- Data Domain File system DDFS (Since 2001);
- Zettabyte File System (ZFS, Oracle) (since 2009);
- B-tree file system (BTRFS) (since August 2014);
- LiveDFS;

- Windows 10 Technical Preview (2016), Microsoft Windows Server 2012/2016 feature of NTFS post-process deduplication;
- OpenDedup (SDFS) in-line deduplication;

Previous works

Focus on:

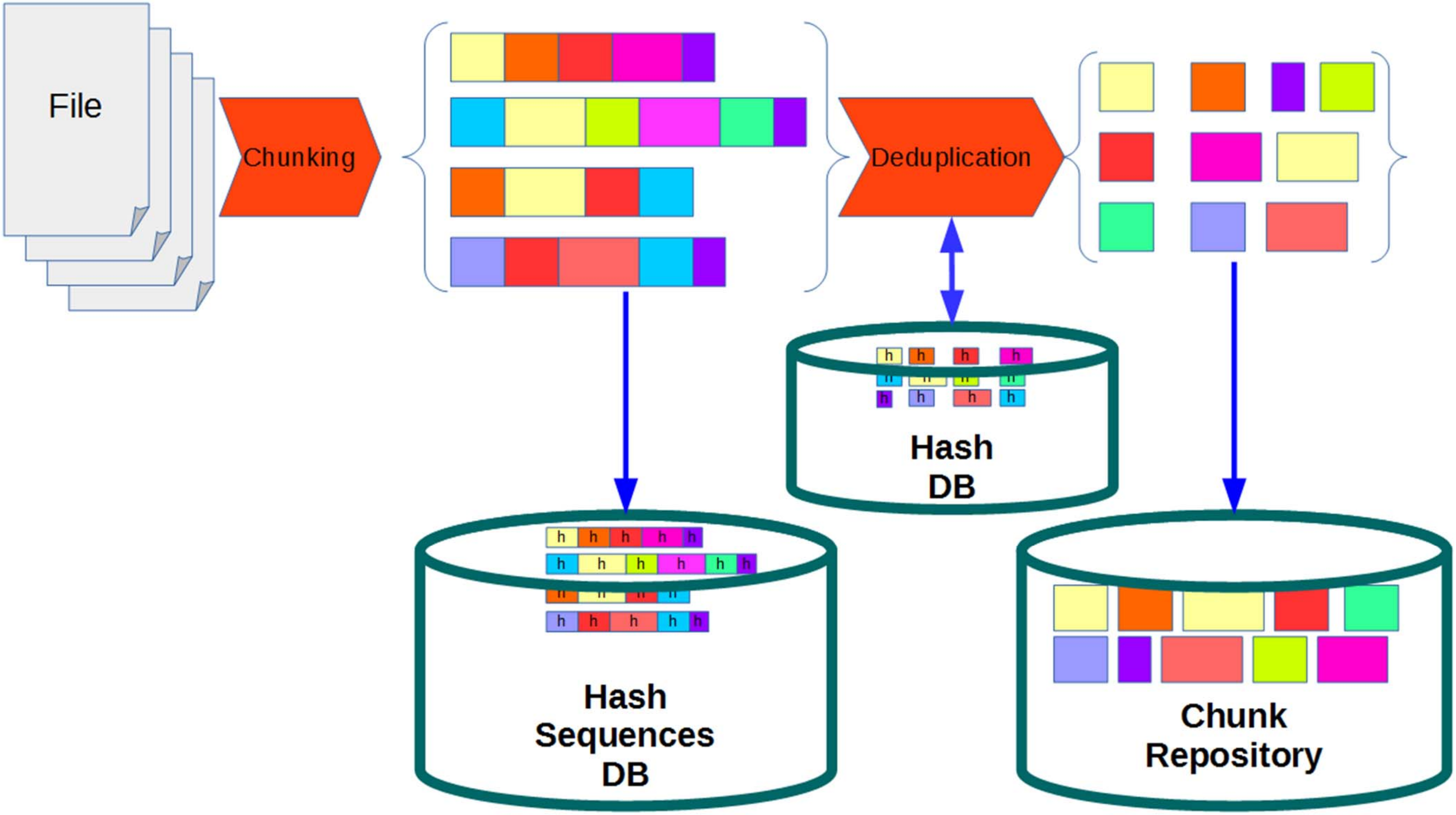
- Deduplication algorithms and deduplication efficiency;
- Deduplication usage - storage technologies to save space to store and analyze forensics data;

Many authors indicate:

  - the need for thorough study using experimental data and physical acquisition;
  - the importance of marker identification, in order to help storage technology recognition;

Lack of studies that explain actual implementations of deduplication from a forensic point of view.

Deduplication – operating principle



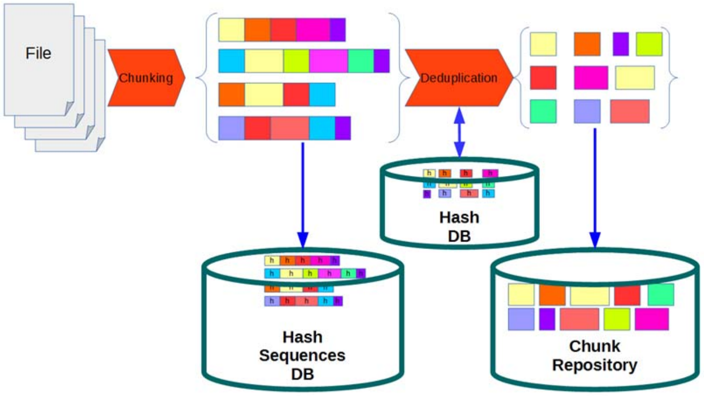
Deduplication – operating principle

- Deduplication transforms a set of files in an organized set of chunks;
- Deduplication computes hash of each chunk and if not jet present it stores hash in a hashdb and it stores chunk in a repository;

- Deduplication stores chunks that are common to multiple files only once;
- Deduplication keeps a sequence of hash for each file;

- Deduplication process can be In-line / Off-line;

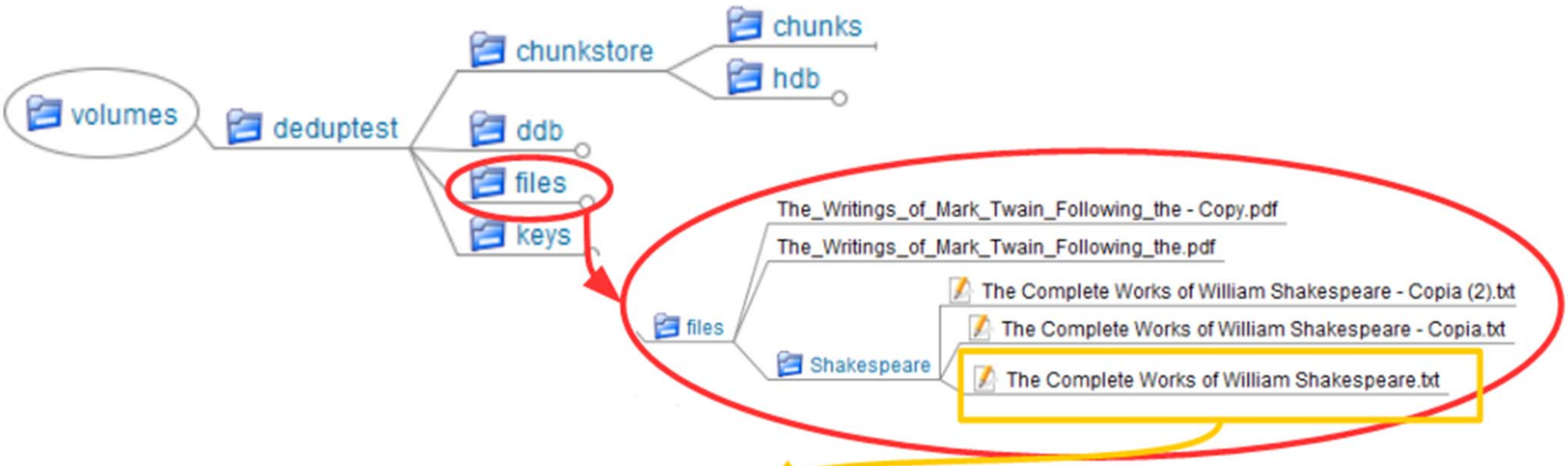
- Chunks generation
  - Fixed Length (Fast)
  - Variable length (Better performance)
    - Rabin algorithm

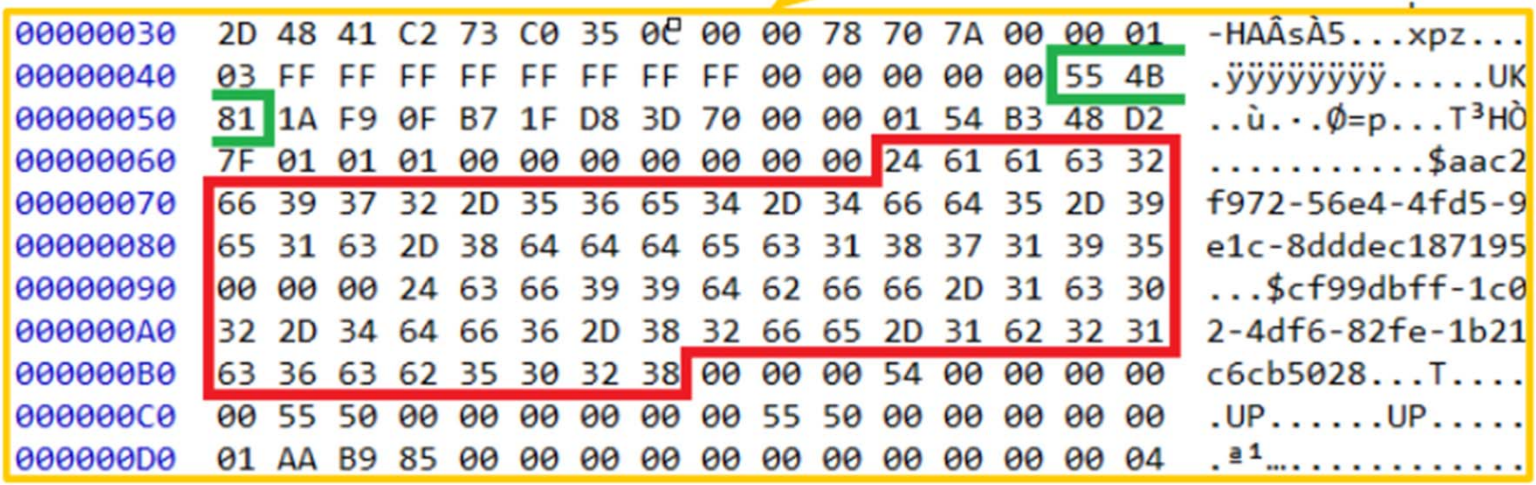


OpenDedup

- OpenDedup
  - OpenDedup (SDFS) inline deduplication;
  - Open source
  - Filesystem in user space (FUSE)
  - Hashes computed using MurMurHash3
  - File Chunking uses Rabin algorithm fingerprint

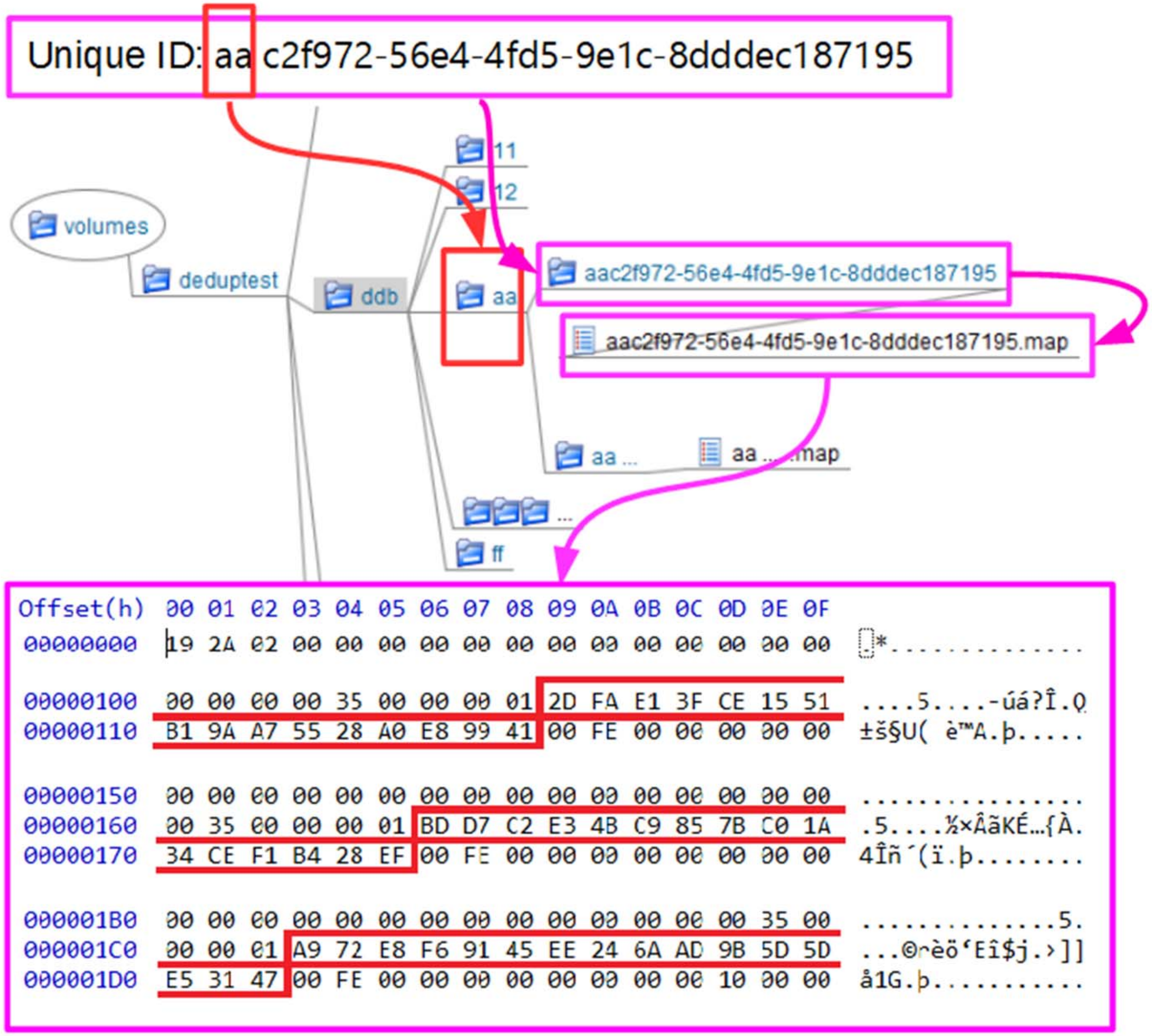
OpenDedup


volumes
deduptest
chunkstore
chunks
hdb
ddb
files
keys
files
The_Writings_of_Mark_Twain_Following_the - Copy.pdf
The_Writings_of_Mark_Twain_Following_the.pdf
Shakespeare
The Complete Works of William Shakespeare - Copia (2).txt
The Complete Works of William Shakespeare - Copia.txt
The Complete Works of William Shakespeare.txt



```
00000030  2D 48 41 C2 73 C0 35 0C 00 00 78 70 7A 00 00 01  -HAÂsÀ5...xpz...
00000040  03 FF FF FF FF FF FF FF FF 00 00 00 00 00 55 4B  .ÿÿÿÿÿÿÿÿ.....UK
00000050  81 1A F9 0F B7 1F D8 3D 70 00 00 01 54 B3 48 D2  ..ù.·.Ø=p...T³HÒ
00000060  7F 01 01 01 00 00 00 00 00 00 00 24 61 61 63 32  ...........$aac2
00000070  66 39 37 32 2D 35 36 65 34 2D 34 66 64 35 2D 39  f972-56e4-4fd5-9
00000080  65 31 63 2D 38 64 64 64 65 63 31 38 37 31 39 35  e1c-8dddec187195
00000090  00 00 00 24 63 66 39 39 64 62 66 66 2D 31 63 30  ...$cf99dbff-1c0
000000A0  32 2D 34 64 66 36 2D 38 32 66 65 2D 31 62 32 31  2-4df6-82fe-1b21
000000B0  63 36 63 62 35 30 32 38 00 00 00 54 00 00 00 00  c6cb5028...T....
000000C0  00 55 50 00 00 00 00 00 00 55 50 00 00 00 00 00  .UP......UP.....
000000D0  01 AA B9 85 00 00 00 00 00 00 00 00 00 00 00 04  .ª¹…............
```

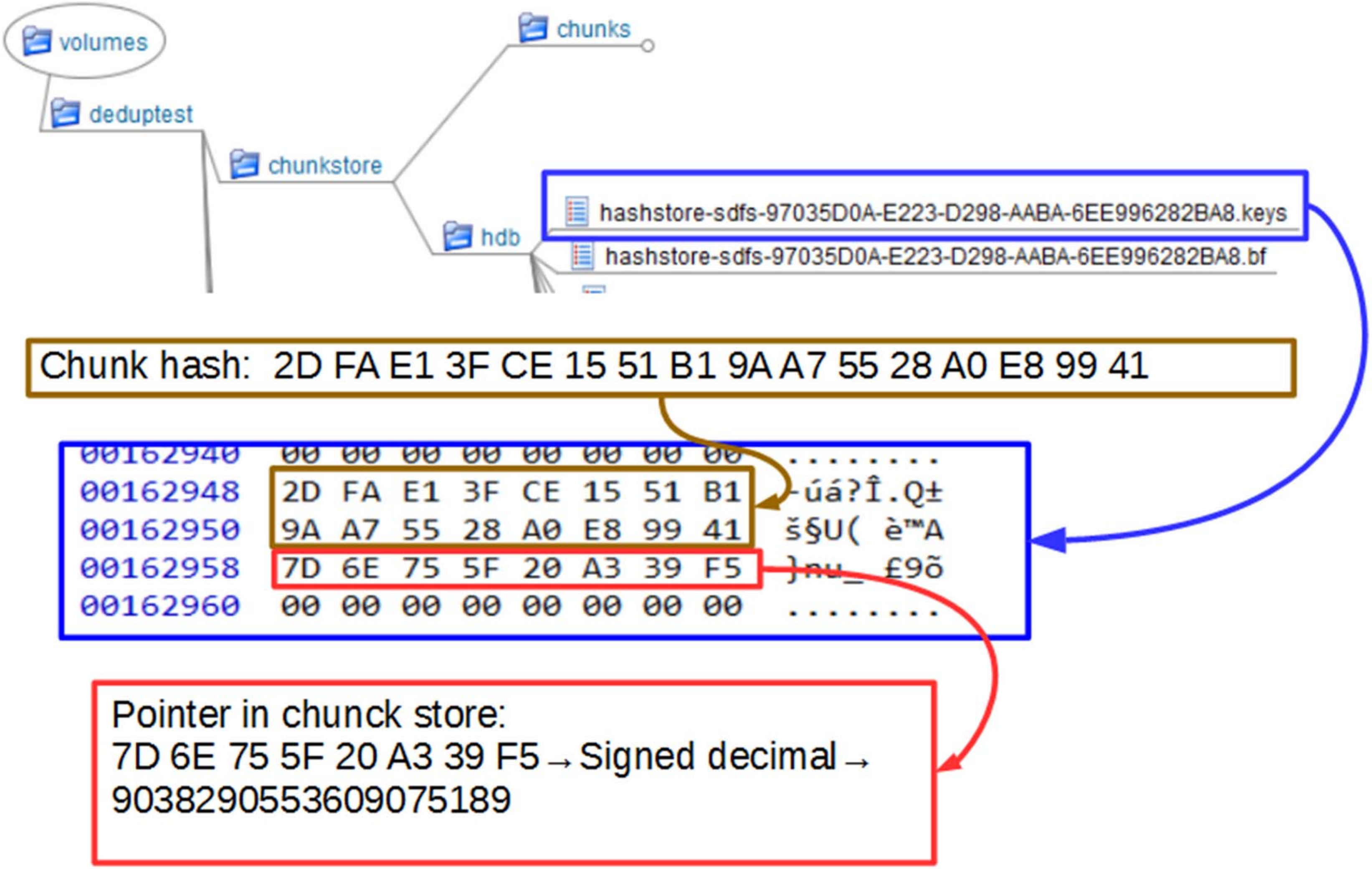
OpenDedup


Unique ID: aa c2f972-56e4-4fd5-9e1c-8dddec187195
volumes
deduptest
ddb
11
12
aa
aac2f972-56e4-4fd5-9e1c-8dddec187195
aac2f972-56e4-4fd5-9e1c-8dddec187195.map
aa ...
aa ... .map
...
ff

```
Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
00000000  19 2A 02 00 00 00 00 00 00 00 00 00 00 00 00 00  □*..............

00000100  00 00 00 00 35 00 00 00 01 2D FA E1 3F CE 15 51  ....5....-úá?Î.Q
00000110  B1 9A A7 55 28 A0 E8 99 41 00 FE 00 00 00 00 00  ±š§U( è™A.þ.....

00000150  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00000160  00 35 00 00 00 01 BD D7 C2 E3 4B C9 85 7B C0 1A  .5....½×ÂãKÉ…{À.
00000170  34 CE F1 B4 28 EF 00 FE 00 00 00 00 00 00 00 00  4Îñ´(ï.þ........

000001B0  00 00 00 00 00 00 00 00 00 00 00 00 00 00 35 00  ..............5.
000001C0  00 00 01 A9 72 E8 F6 91 45 EE 24 6A AD 9B 5D 5D  ...©rèö‘Eî$j.›]]
000001D0  E5 31 47 00 FE 00 00 00 00 00 00 00 00 10 00 00  å1G.þ...........
```

The .map file contains the hash sequence

OpenDedup

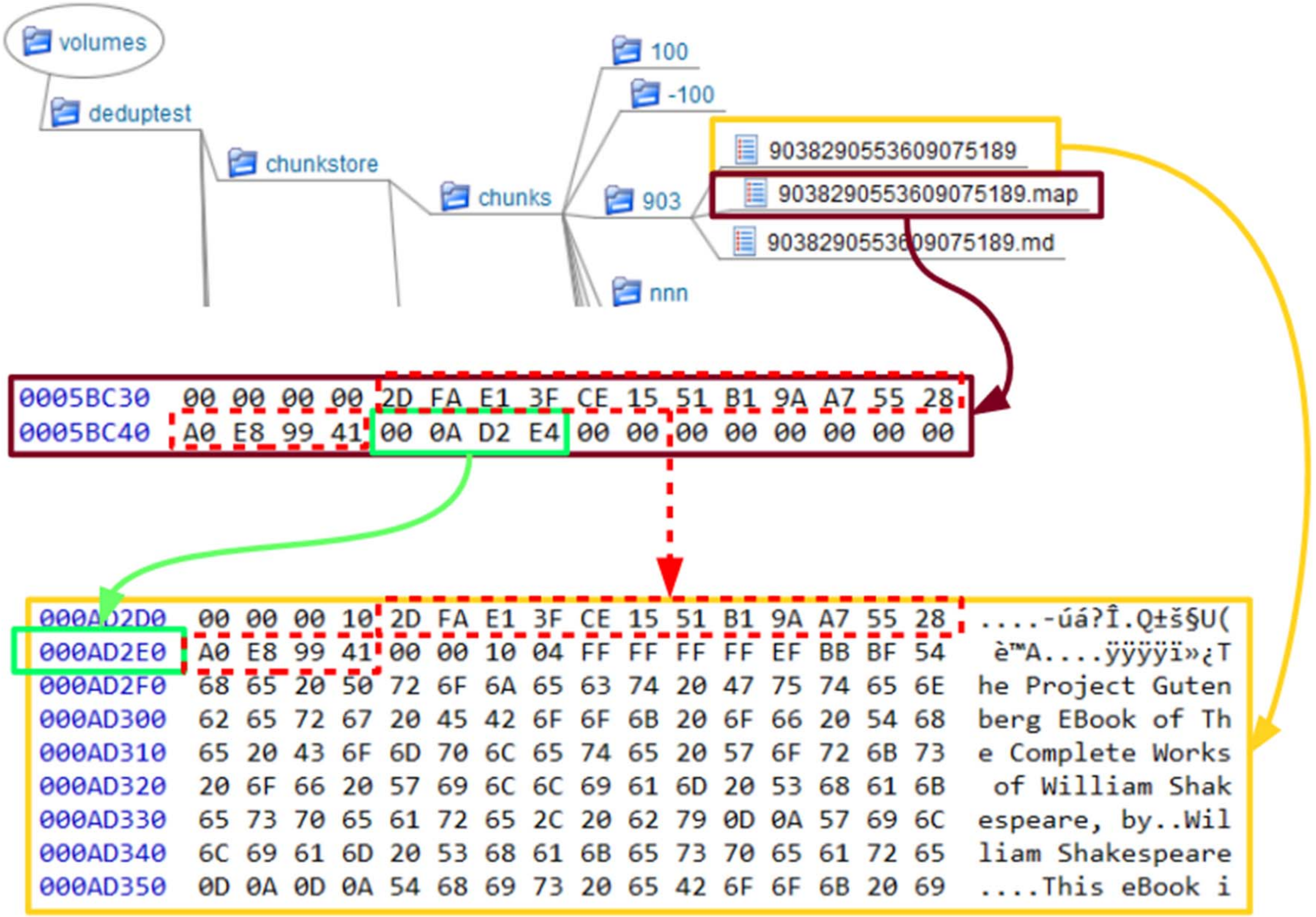

volumes
deduptest
chunkstore
chunks
hdb
hashstore-sdfs-97035D0A-E223-D298-AABA-6EE996282BA8.keys
hashstore-sdfs-97035D0A-E223-D298-AABA-6EE996282BA8.bf

Chunk hash: 2D FA E1 3F CE 15 51 B1 9A A7 55 28 A0 E8 99 41

```
00162940  00 00 00 00 00 00 00 00  ........
00162948  2D FA E1 3F CE 15 51 B1  -úá?Î.Q±
00162950  9A A7 55 28 A0 E8 99 41  š§U( è™A
00162958  7D 6E 75 5F 20 A3 39 F5  }nu_ £9õ
00162960  00 00 00 00 00 00 00 00  ........
```

Pointer in chunck store:
7D 6E 75 5F 20 A3 39 F5 → Signed decimal →
9038290553609075189

OpenDedup



```
0005BC30  00 00 00 00 2D FA E1 3F CE 15 51 B1 9A A7 55 28
0005BC40  A0 E8 99 41 00 0A D2 E4 00 00 00 00 00 00 00 00
```

```
000AD2D0  00 00 00 10 2D FA E1 3F CE 15 51 B1 9A A7 55 28  ....-úá?Î.Q±š§U(
000AD2E0  A0 E8 99 41 00 00 10 04 FF FF FF FF EF BB BF 54   è™A....ÿÿÿÿï»¿T
000AD2F0  68 65 20 50 72 6F 6A 65 63 74 20 47 75 74 65 6E  he Project Guten
000AD300  62 65 72 67 20 45 42 6F 6F 6B 20 6F 66 20 54 68  berg EBook of Th
000AD310  65 20 43 6F 6D 70 6C 65 74 65 20 57 6F 72 6B 73  e Complete Works
000AD320  20 6F 66 20 57 69 6C 6C 69 61 6D 20 53 68 61 6B   of William Shak
000AD330  65 73 70 65 61 72 65 2C 20 62 79 0D 0A 57 69 6C  espeare, by..Wil
000AD340  6C 69 61 6D 20 53 68 61 6B 65 73 70 65 61 72 65  liam Shakespeare
000AD350  0D 0A 0D 0A 54 68 69 73 20 65 42 6F 6F 6B 20 69  ....This eBook i
```

Microsoft Windows Deduplication

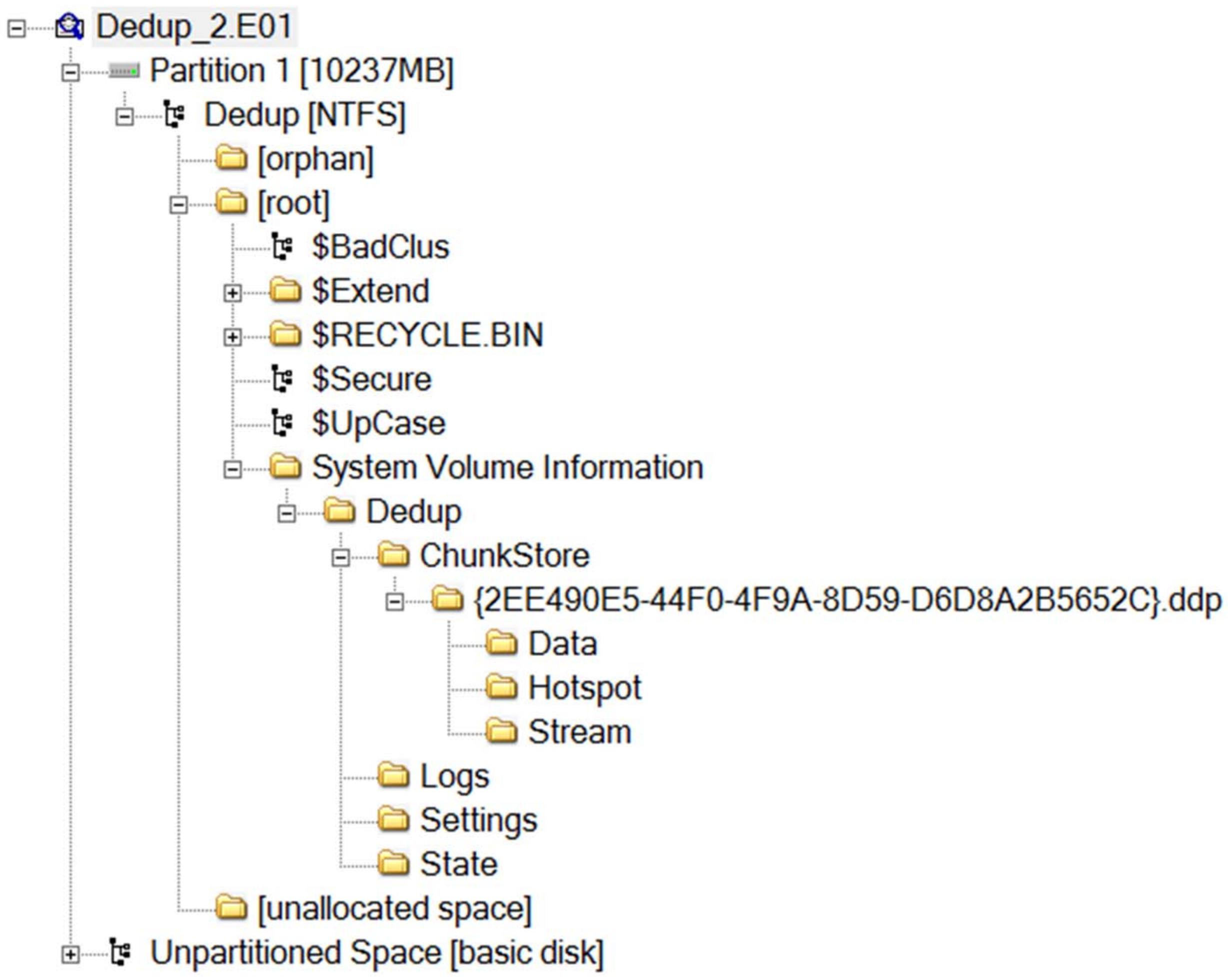
- Integrated in NTFS
  - uses the $MFT, attribute Reparse_Point
  - Leave traces in journal log

- Off-line process (age, usage, file type)
  - The process leaves artifacts

- Chunk are compressed
  - High entropy on devices (all chunks contains compressed data)

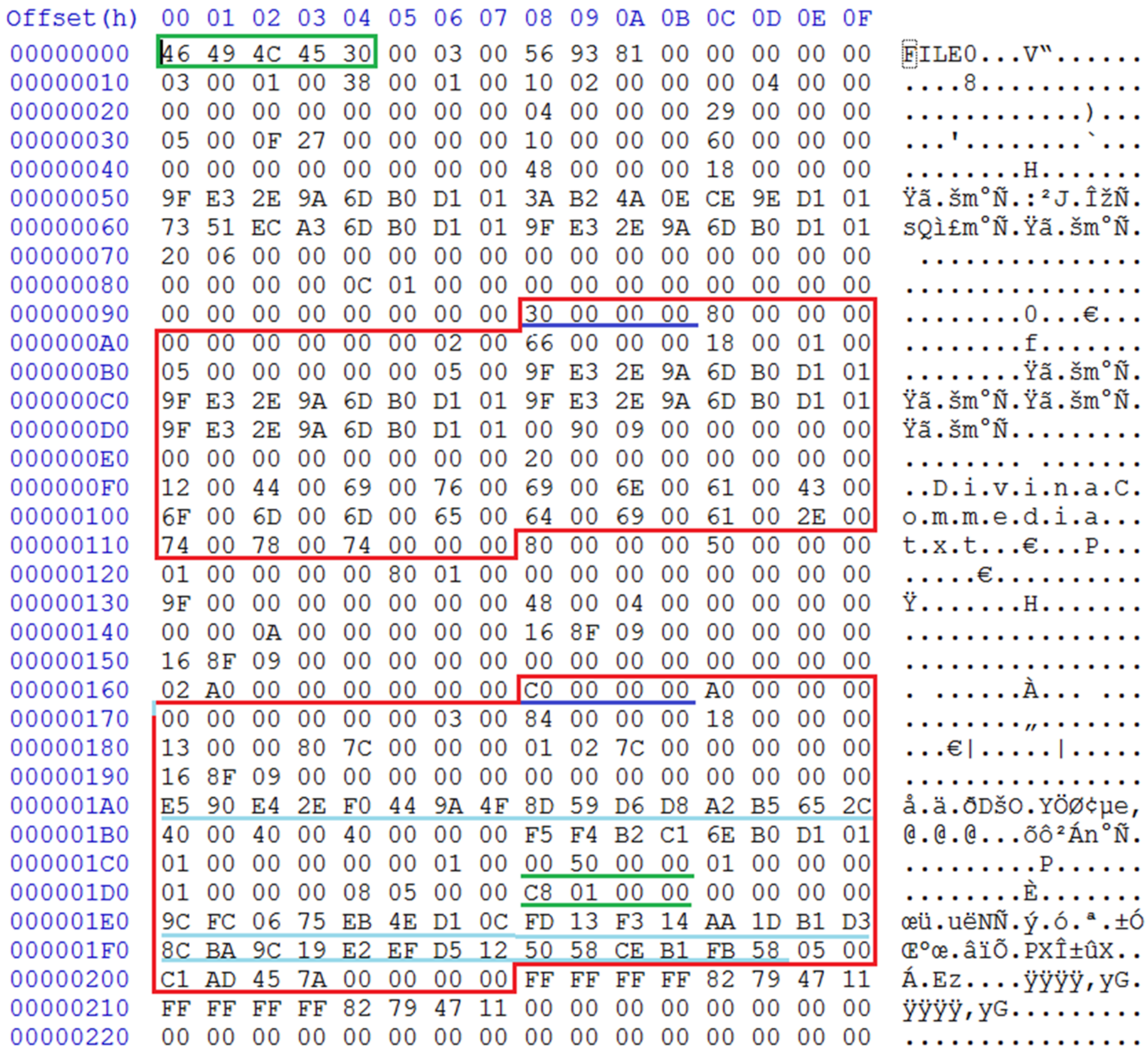
- System Volume Information contains the chunkstore structure
  - the stream container
  - the data chunk container
  - the hotspot container

Microsoft Windows Deduplication



```
Dedup_2.E01
    Partition 1 [10237MB]
        Dedup [NTFS]
            [orphan]
            [root]
                $BadClus
                $Extend
                $RECYCLE.BIN
                $Secure
                $UpCase
                System Volume Information
                    Dedup
                        ChunkStore
                            {2EE490E5-44F0-4F9A-8D59-D6D8A2B5652C}.ddp
                                Data
                                Hotspot
                                Stream
                        Logs
                        Settings
                        State
            [unallocated space]
    Unpartitioned Space [basic disk]
```

$MFT Record



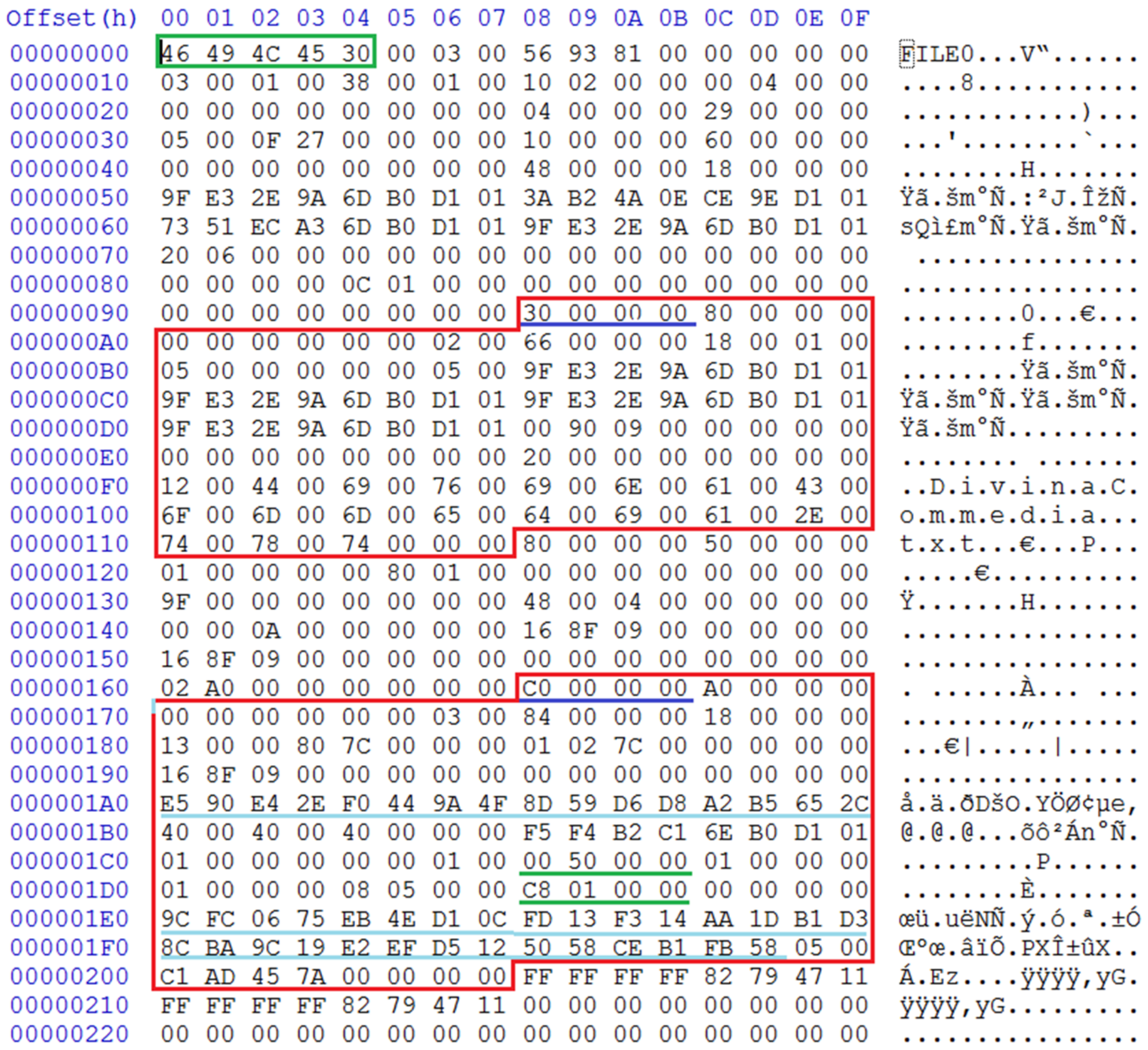
```
Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00000000  46 49 4C 45 30 00 03 00 56 93 81 00 00 00 00 00  FILE0...V“......
00000010  03 00 01 00 38 00 01 00 10 02 00 00 00 04 00 00  ....8...........
00000020  00 00 00 00 00 00 00 00 04 00 00 00 29 00 00 00  ............)...
00000030  05 00 0F 27 00 00 00 00 10 00 00 00 60 00 00 00  ...'........`...
00000040  00 00 00 00 00 00 00 00 48 00 00 00 18 00 00 00  ........H.......
00000050  9F E3 2E 9A 6D B0 D1 01 3A B2 4A 0E CE 9E D1 01  Ÿã.šm°Ñ.:²J.ÎžÑ.
00000060  73 51 EC A3 6D B0 D1 01 9F E3 2E 9A 6D B0 D1 01  sQì£m°Ñ.Ÿã.šm°Ñ.
00000070  20 06 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ...............
00000080  00 00 00 00 0C 01 00 00 00 00 00 00 00 00 00 00  ................
00000090  00 00 00 00 00 00 00 00 30 00 00 00 80 00 00 00  ........0...€...
000000A0  00 00 00 00 00 00 02 00 66 00 00 00 18 00 01 00  ........f.......
000000B0  05 00 00 00 00 00 05 00 9F E3 2E 9A 6D B0 D1 01  ........Ÿã.šm°Ñ.
000000C0  9F E3 2E 9A 6D B0 D1 01 9F E3 2E 9A 6D B0 D1 01  Ÿã.šm°Ñ.Ÿã.šm°Ñ.
000000D0  9F E3 2E 9A 6D B0 D1 01 00 90 09 00 00 00 00 00  Ÿã.šm°Ñ.........
000000E0  00 00 00 00 00 00 00 00 20 00 00 00 00 00 00 00  ........ .......
000000F0  12 00 44 00 69 00 76 00 69 00 6E 00 61 00 43 00  ..D.i.v.i.n.a.C.
00000100  6F 00 6D 00 6D 00 65 00 64 00 69 00 61 00 2E 00  o.m.m.e.d.i.a...
00000110  74 00 78 00 74 00 00 00 80 00 00 00 50 00 00 00  t.x.t...€...P...
00000120  01 00 00 00 00 80 01 00 00 00 00 00 00 00 00 00  .....€..........
00000130  9F 00 00 00 00 00 00 00 48 00 04 00 00 00 00 00  Ÿ.......H.......
00000140  00 00 0A 00 00 00 00 00 16 8F 09 00 00 00 00 00  ................
00000150  16 8F 09 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00000160  02 A0 00 00 00 00 00 00 C0 00 00 00 A0 00 00 00  . ......À... ...
00000170  00 00 00 00 00 00 03 00 84 00 00 00 18 00 00 00  ........„.......
00000180  13 00 00 80 7C 00 00 00 01 02 7C 00 00 00 00 00  ...€|.....|.....
00000190  16 8F 09 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
000001A0  E5 90 E4 2E F0 44 9A 4F 8D 59 D6 D8 A2 B5 65 2C  å.ä.ðDšO.YÖØ¢µe,
000001B0  40 00 40 00 40 00 00 00 F5 F4 B2 C1 6E B0 D1 01  @.@.@...õô²Án°Ñ.
000001C0  01 00 00 00 00 00 01 00 00 50 00 00 01 00 00 00  .........P......
000001D0  01 00 00 00 08 05 00 00 C8 01 00 00 00 00 00 00  ........È.......
000001E0  9C FC 06 75 EB 4E D1 0C FD 13 F3 14 AA 1D B1 D3  œü.uëNÑ.ý.ó.ª.±Ó
000001F0  8C BA 9C 19 E2 EF D5 12 50 58 CE B1 FB 58 05 00  Œºœ.âïÕ.PXÎ±ûX..
00000200  C1 AD 45 7A 00 00 00 00 FF FF FF FF 82 79 47 11  Á.Ez....ÿÿÿÿ‚yG.
00000210  FF FF FF FF 82 79 47 11 00 00 00 00 00 00 00 00  ÿÿÿÿ‚yG.........
00000220  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
```

- 0x000000C0 $REPARSE_POINT Attribute
- Offset 0x01A0 (E5 90 E4 2E - F0 44 - 9A 4F - 8D 59 - D6 D8 A2 B5 65 2C)

  → identifies the ChunkStore

  →{2EE490E5- 44F0-4F9A- 8D59-D6D8A2B5652C}.ddp

$MFT Record



```
Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00000000  46 49 4C 45 30 00 03 00 56 93 81 00 00 00 00 00  FILE0...V“......
00000010  03 00 01 00 38 00 01 00 10 02 00 00 00 04 00 00  ....8...........
00000020  00 00 00 00 00 00 00 00 04 00 00 00 29 00 00 00  ............)...
00000030  05 00 0F 27 00 00 00 00 10 00 00 00 60 00 00 00  ...'........`...
00000040  00 00 00 00 00 00 00 00 48 00 00 00 18 00 00 00  ........H.......
00000050  9F E3 2E 9A 6D B0 D1 01 3A B2 4A 0E CE 9E D1 01  Ÿã.šm°Ñ.:²J.ÎžÑ.
00000060  73 51 EC A3 6D B0 D1 01 9F E3 2E 9A 6D B0 D1 01  sQì£m°Ñ.Ÿã.šm°Ñ.
00000070  20 06 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ...............
00000080  00 00 00 00 0C 01 00 00 00 00 00 00 00 00 00 00  ................
00000090  00 00 00 00 00 00 00 00 30 00 00 00 80 00 00 00  ........0...€...
000000A0  00 00 00 00 00 00 02 00 66 00 00 00 18 00 01 00  ........f.......
000000B0  05 00 00 00 00 00 05 00 9F E3 2E 9A 6D B0 D1 01  ........Ÿã.šm°Ñ.
000000C0  9F E3 2E 9A 6D B0 D1 01 9F E3 2E 9A 6D B0 D1 01  Ÿã.šm°Ñ.Ÿã.šm°Ñ.
000000D0  9F E3 2E 9A 6D B0 D1 01 00 90 09 00 00 00 00 00  Ÿã.šm°Ñ.........
000000E0  00 00 00 00 00 00 00 00 20 00 00 00 00 00 00 00  ........ .......
000000F0  12 00 44 00 69 00 76 00 69 00 6E 00 61 00 43 00  ..D.i.v.i.n.a.C.
00000100  6F 00 6D 00 6D 00 65 00 64 00 69 00 61 00 2E 00  o.m.m.e.d.i.a...
00000110  74 00 78 00 74 00 00 00 80 00 00 00 50 00 00 00  t.x.t...€...P...
00000120  01 00 00 00 00 80 01 00 00 00 00 00 00 00 00 00  .....€..........
00000130  9F 00 00 00 00 00 00 00 48 00 04 00 00 00 00 00  Ÿ.......H.......
00000140  00 00 0A 00 00 00 00 00 16 8F 09 00 00 00 00 00  ................
00000150  16 8F 09 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00000160  02 A0 00 00 00 00 00 00 C0 00 00 00 A0 00 00 00  . ......À... ...
00000170  00 00 00 00 00 00 03 00 84 00 00 00 18 00 00 00  ........„.......
00000180  13 00 00 80 7C 00 00 00 01 02 7C 00 00 00 00 00  ...€|.....|.....
00000190  16 8F 09 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
000001A0  E5 90 E4 2E F0 44 9A 4F 8D 59 D6 D8 A2 B5 65 2C  å.ä.ðDšO.YÖØ¢µe,
000001B0  40 00 40 00 40 00 00 00 F5 F4 B2 C1 6E B0 D1 01  @.@.@...õô²Án°Ñ.
000001C0  01 00 00 00 00 00 01 00 00 50 00 00 01 00 00 00  .........P......
000001D0  01 00 00 00 08 05 00 00 C8 01 00 00 00 00 00 00  ........È.......
000001E0  9C FC 06 75 EB 4E D1 0C FD 13 F3 14 AA 1D B1 D3  œü.uëNÑ.ý.ó.ª.±Ó
000001F0  8C BA 9C 19 E2 EF D5 12 50 58 CE B1 FB 58 05 00  Œ°œ.âïÕ.PXÎ±ûX..
00000200  C1 AD 45 7A 00 00 00 00 FF FF FF FF 82 79 47 11  Á.Ez....ÿÿÿÿ,yG.
00000210  FF FF FF FF 82 79 47 11 00 00 00 00 00 00 00 00  ÿÿÿÿ,yG.........
00000220  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
```

- Offset 0x01E0
  9C FC 06 75 EB 4E D1 0C
  FD 13 F3 14 AA 1D B1
  D3 8C BA 9C 19 E2 EF
  D5 12 50 58 CE B1 FB 58
  → identifies the hash sequence in the stream

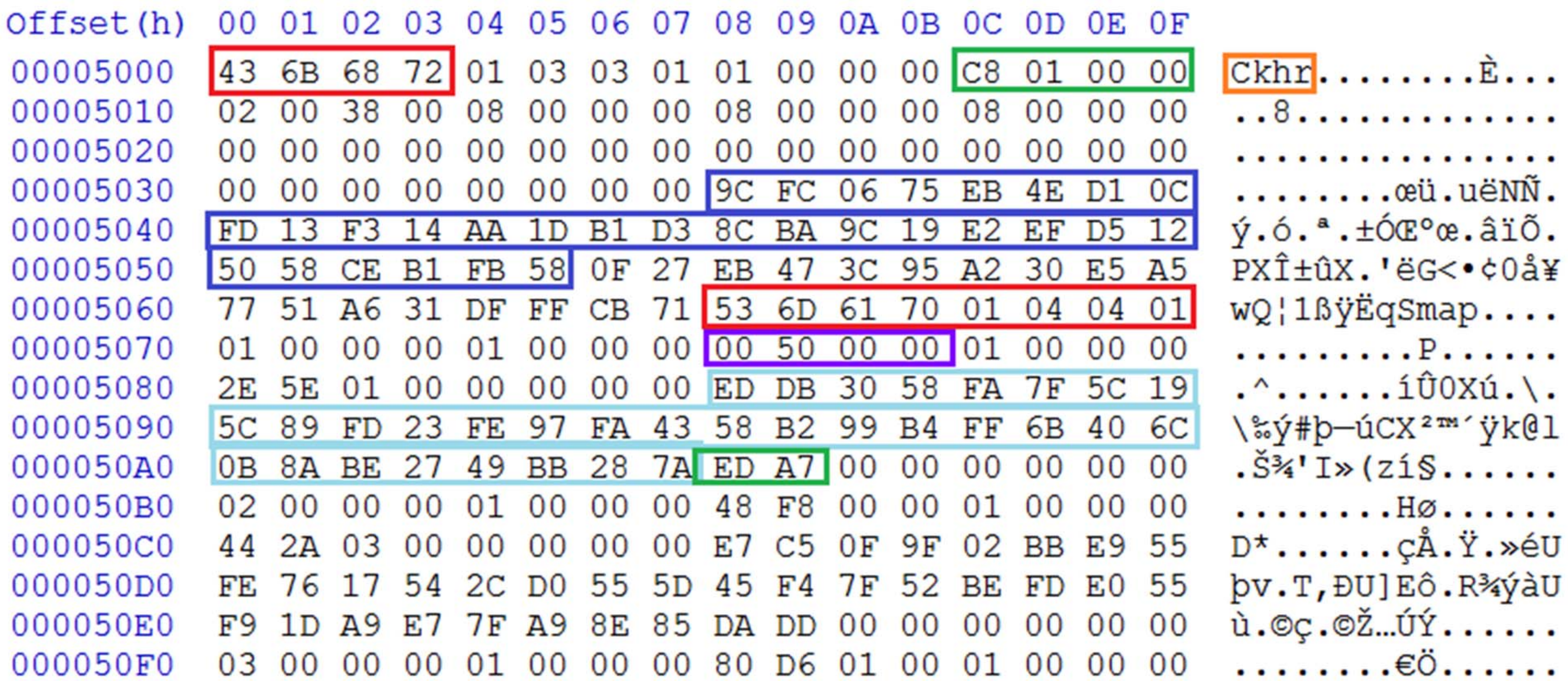
- Offset 0x01C8
  00 50 → 0x5000
  sequence start address

- Offset 0x01D8
  C8 01→ 0x01C8
  Sequence length

$MFT Record → Stream

- From $MFT analysis we know
  - ChunkStore ID:
    {2EE490E5- 44F0- 4F9A- 8D59- D6D8A2B5652C}.ddp
  - Stream sequence id:
    9C FC 06 75 EB 4E D1 0C FD 13 F3 14 AA 1D B1 D3 8C BA 9C 19 E2 EF D5 12 50 58 CE B1 FB 58
  - Sequence start address:
    0x5000
  - Sequence length:
    0x01C8

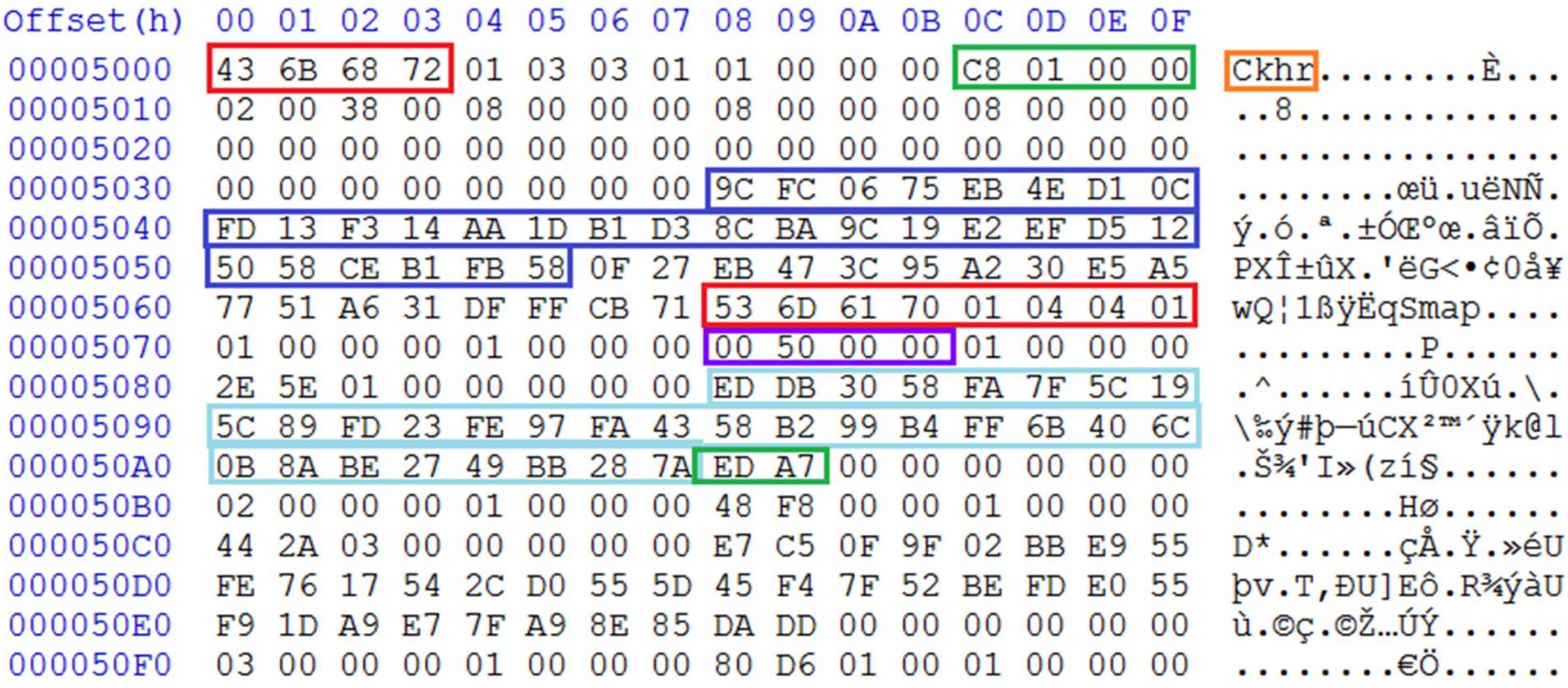
Stream Container



```
Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00005000  43 6B 68 72 01 03 03 01 01 00 00 00 C8 01 00 00  Ckhr........È...
00005010  02 00 38 00 08 00 00 00 08 00 00 00 08 00 00 00  ..8.............
00005020  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00005030  00 00 00 00 00 00 00 00 9C FC 06 75 EB 4E D1 0C  ........œü.uëNÑ.
00005040  FD 13 F3 14 AA 1D B1 D3 8C BA 9C 19 E2 EF D5 12  ý.ó.ª.±ÓŒ°œ.âïÕ.
00005050  50 58 CE B1 FB 58 0F 27 EB 47 3C 95 A2 30 E5 A5  PXÎ±ûX.'ëG<•¢0å¥
00005060  77 51 A6 31 DF FF CB 71 53 6D 61 70 01 04 04 01  wQ¦1ßÿËqSmap....
00005070  01 00 00 00 01 00 00 00 00 50 00 00 01 00 00 00  .........P......
00005080  2E 5E 01 00 00 00 00 00 ED DB 30 58 FA 7F 5C 19  .^......íÛ0Xú.\.
00005090  5C 89 FD 23 FE 97 FA 43 58 B2 99 B4 FF 6B 40 6C  \‰ý#þ—úCX²™´ÿk@l
000050A0  0B 8A BE 27 49 BB 28 7A ED A7 00 00 00 00 00 00  .Š¾'I»(zí§......
000050B0  02 00 00 00 01 00 00 00 48 F8 00 00 01 00 00 00  ........Hø......
000050C0  44 2A 03 00 00 00 00 00 E7 C5 0F 9F 02 BB E9 55  D*......çÅ.Ÿ.»éU
000050D0  FE 76 17 54 2C D0 55 5D 45 F4 7F 52 BE FD E0 55  þv.T,ÐU]Eô.R¾ýàU
000050E0  F9 1D A9 E7 7F A9 8E 85 DA DD 00 00 00 00 00 00  ù.©ç.©Ž…ÚÝ......
000050F0  03 00 00 00 01 00 00 00 80 D6 01 00 01 00 00 00  ........€Ö......
```

- Displacement 0x00 →Ckhr – Marker of sequence
- Displacement 0x0C → Sequence length
- Displacement 0x30 →Stream sequence id

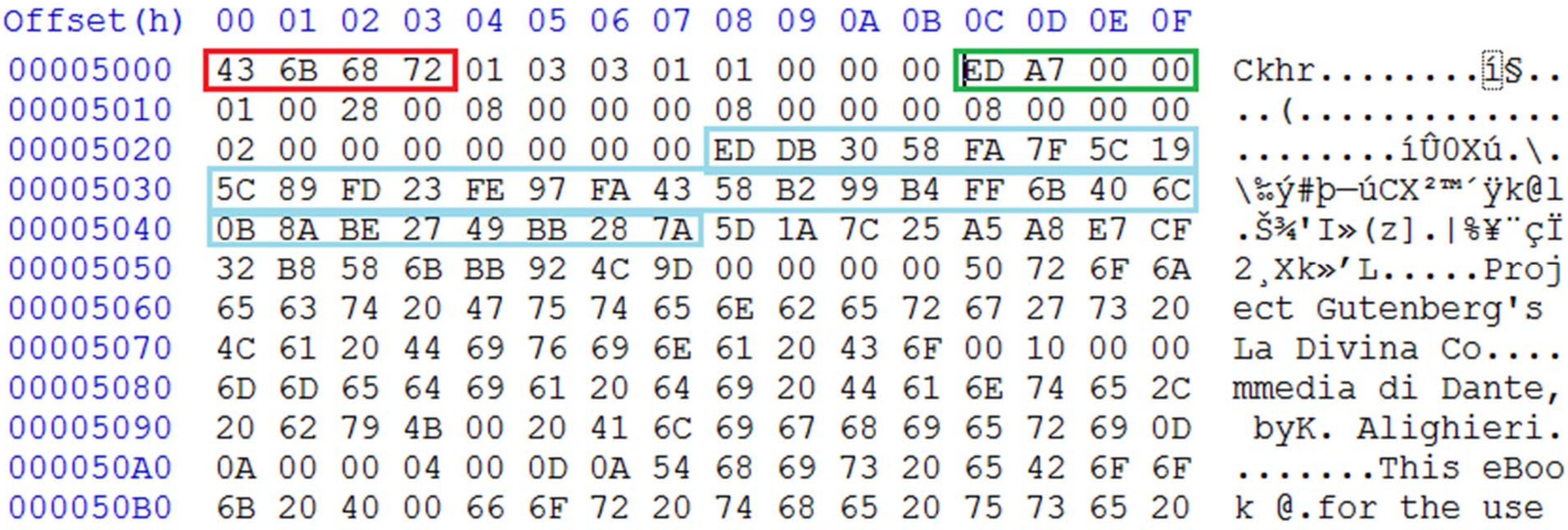
Stream Container



```
Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00005000  43 6B 68 72 01 03 03 01 01 00 00 00 C8 01 00 00  Ckhr........È...
00005010  02 00 38 00 08 00 00 00 08 00 00 00 08 00 00 00  ..8.............
00005020  00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00005030  00 00 00 00 00 00 00 00 9C FC 06 75 EB 4E D1 0C  ........œü.uëNÑ.
00005040  FD 13 F3 14 AA 1D B1 D3 8C BA 9C 19 E2 EF D5 12  ý.ó.ª.±ÓŒºœ.âïÕ.
00005050  50 58 CE B1 FB 58 0F 27 EB 47 3C 95 A2 30 E5 A5  PXÎ±ûX.'ëG<•¢0å¥
00005060  77 51 A6 31 DF FF CB 71 53 6D 61 70 01 04 04 01  wQ¦1ßÿËqSmap....
00005070  01 00 00 00 01 00 00 00 00 50 00 00 01 00 00 00  .........P......
00005080  2E 5E 01 00 00 00 00 00 ED DB 30 58 FA 7F 5C 19  .^......íÛ0Xú.\.
00005090  5C 89 FD 23 FE 97 FA 43 58 B2 99 B4 FF 6B 40 6C  \‰ý#þ—úCX²™´ÿk@l
000050A0  0B 8A BE 27 49 BB 28 7A ED A7 00 00 00 00 00 00  .Š¾'I»(zí§......
000050B0  02 00 00 00 01 00 00 00 48 F8 00 00 01 00 00 00  ........Hø......
000050C0  44 2A 03 00 00 00 00 00 E7 C5 0F 9F 02 BB E9 55  D*......çÅ.Ÿ.»éU
000050D0  FE 76 17 54 2C D0 55 5D 45 F4 7F 52 BE FD E0 55  þv.T,ÐU]Eô.R¾ýàU
000050E0  F9 1D A9 E7 7F A9 8E 85 DA DD 00 00 00 00 00 00  ù.©ç.©Ž…ÚÝ......
000050F0  03 00 00 00 01 00 00 00 80 D6 01 00 01 00 00 00  ........€Ö......
```

- Displacement 0x68 →Smap…. – Marker of first hash value
- Displacement 0x78 →Chunk address in chunk container
- Displacement 0x88 →First chunk hash
- Displacement 0xA8 →First chunk length

Chunk Container



```
Offset(h) 00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00005000  43 6B 68 72 01 03 03 01 01 00 00 00 ED A7 00 00  Ckhr........í§..
00005010  01 00 28 00 08 00 00 00 08 00 00 00 08 00 00 00  ..(.............
00005020  02 00 00 00 00 00 00 00 ED DB 30 58 FA 7F 5C 19  ........íÛ0Xú.\.
00005030  5C 89 FD 23 FE 97 FA 43 58 B2 99 B4 FF 6B 40 6C  \‰ý#þ—úCX²™´ÿk@l
00005040  0B 8A BE 27 49 BB 28 7A 5D 1A 7C 25 A5 A8 E7 CF  .Š¾'I»(z].|%¥¨çÏ
00005050  32 B8 58 6B BB 92 4C 9D 00 00 00 00 50 72 6F 6A  2¸Xk»’L.....Proj
00005060  65 63 74 20 47 75 74 65 6E 62 65 72 67 27 73 20  ect Gutenberg's 
00005070  4C 61 20 44 69 76 69 6E 61 20 43 6F 00 10 00 00  La Divina Co....
00005080  6D 6D 65 64 69 61 20 64 69 20 44 61 6E 74 65 2C  mmedia di Dante,
00005090  20 62 79 4B 00 20 41 6C 69 67 68 69 65 72 69 0D   byK. Alighieri.
000050A0  0A 00 00 04 00 0D 0A 54 68 69 73 20 65 42 6F 6F  .......This eBoo
000050B0  6B 20 40 00 66 6F 72 20 74 68 65 20 75 73 65 20  k @.for the use 
```

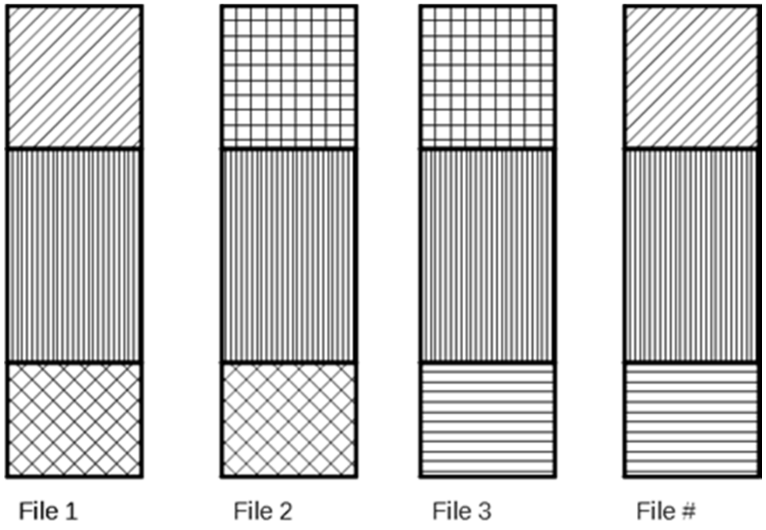
- Displacement 0x00 →Ckhr – Marker of Chunk
- Displacement 0x0C → Chunk length
- Displacement 0x20 → Chunk hash

Operation artifacts

- File deleting
  - The $MFT entry is removed
  - The stream map and chunkstore remain unchanged.
  - File is recoverable immediately after its deletion.
- Optimize process
  - no effect on deleted file stream map and chunk store
- Garbage collection (GC)
  - a regular GC invalidates elements in chunkstore
  - a full GC eliminates all traces (write a new file, leave artifacts of the old one)
  - GC process leaves artifacts in unallocated space of the volume

Importance of hash sequence

- Without the knowledge of chunks concatenation sequence, is impossible to do an accurate reconstruction work;
- Hash sequence is the only way to be sure of the accuracy of file reconstruction;
- The Rabin algorithm uses the output of a polynomial function, and cut the files where a fixed fingerprint is present.



- File 1 – Original file, lenght equal to 3 chunks
- File 2 – File 1 first lines modified
- File 3 – File 2 last lines modified
- File # - File created concatenating 3 chunks recovered from chunk repository. This is a valid file, but never existed in the file system

Conclusion

- Topics covered
  - Analysis of deduplicated file systems;
  - Identification of the elements of the file systems;
  - Recovering deduplicated file system manually;
  - Traces left from deduplication process;
  - Traces left after file deletion;
  - Importance of hash sequence;

Thank you

Thank you