MD5 Checksums

Keith Merrington

APP06

MD5 Checksums

- Some History
- What is a MD5 Checksum
- The MD5 algorithm
- Where it is used
- How to use it
- OS/2 MD5 programs
- Other Programs using MD5

Some Background History

- Checksums originated in the need to verify that data was transferred correctly as hardware was not reliable.

- Examples:
  - Parity bit
  - CRC (Cyclic Redundancy Check)

The MD5 Checksum

- MD5 (Message-Digest algorithm 5) is one in a series of message digest algorithms designed by Prof. R. Rivest in 1994

- MD5 was designed in 1991 to be a secure replacement for MD4, which replaced MD3, etc.

The MD5 Checksum

- A MD5 Checksum uses a 128 bit hash value.

- The MD5 checksum of

“The quick brown fox jumps over the lazy dog"

is

9e107d9d372bb6826bd81d3542a419d6

The MD5 Checksum

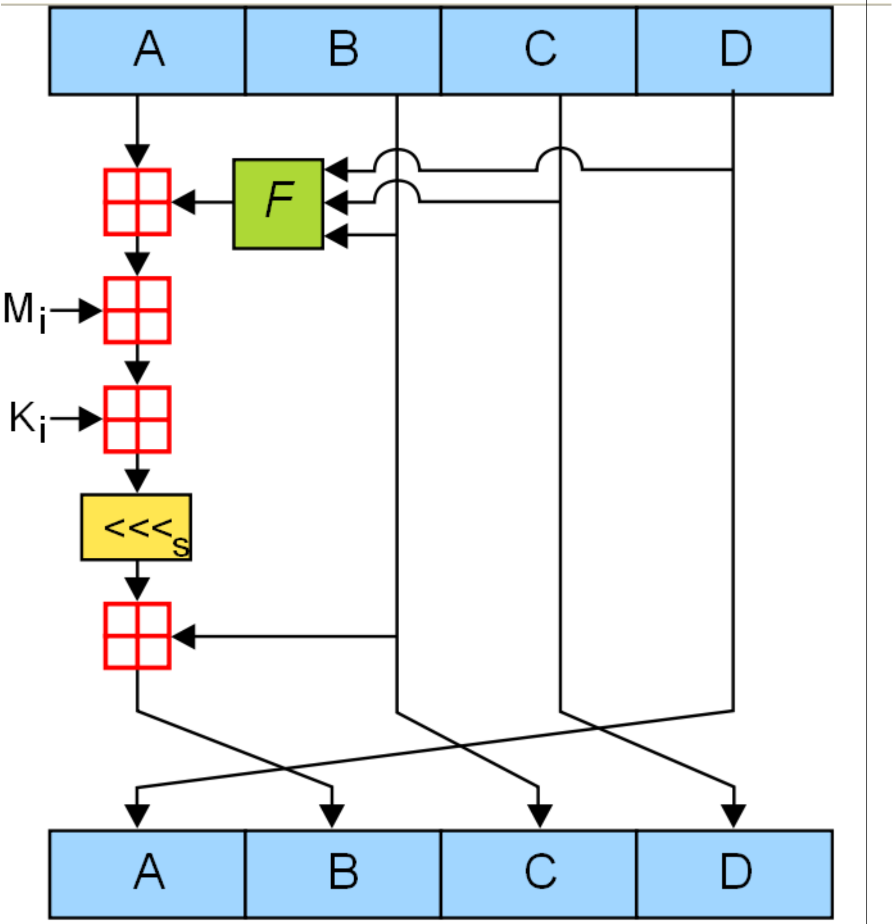
- The MD5 algorithm processes a variable-length message into a fixed-length output of 128 bits.
- The input message is broken up into chunks of 512-bit blocks. This is achieved by padding the input message so that its length is divisible by 512.
- Then each block is broken up into blocks of 32 bytes and then added and ored in a particular way depending on which 32 byte block is now selected.

The MD5 Checksum

One MD5 operation. MD5 consists of 64 of these operations, grouped in four rounds of 16 operations. *F* is a nonlinear function; one function is used in each round. *Mi* denotes a 32-bit block of the message input, and *Ki* denotes a 32-bit constant, different for each operation.

There are four possible functions *F*; a different one is used in each round:

F(X,Y,Z)=(X&&Y)||(!X&&Z)
G(X,Y,Z)=(X&&Z)||(Y&&!Z)
H(X,Y,Z)=X^^Y^^Z
I (X,Y,Z)=Y^^(X||!Z)



Collisions

- This is the term given to the possibility of two different pieces of data producing the same checksum!
- In 1995, collisions were found both in MD5 and SHA1.
- In 2005, researchers were able to create pairs of PostScript documents and X.509 certificates with the same MD5 hash.

Collisions

- Until now only documents having different lengths have been found to give the same MD5 checksum.
- For data integrity checks this makes no real difference!
- For validation (certificates) it will.

Where it is used

- MD5 algorithm is currently used to check that pieces of data are identical.
  - A file transferred across the net arrives correctly
  - A file burnt on a CD is correct
- To create difference files for very large files.

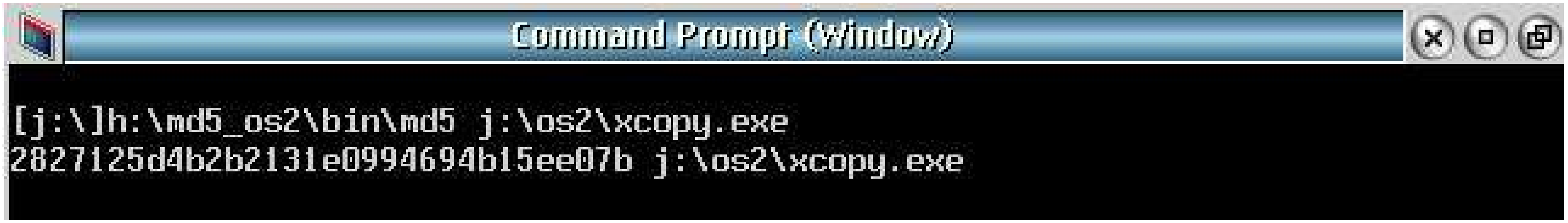
Why use MD5

- The MD5 checksum is platform independent.
- A variety of programs are available to generate and check MD5 checksums on various platforms.
- It was designed to be fast on 32-bit machines

How to use

- In order to check a checksum two items are required
  - The checksum of ‘the file’
  - The item to be checked (filename)

Example :MD5 filename



```
[j:\]h:\md5_os2\bin\md5 j:\os2\xcopy.exe
2827125d4b2b2131e0994694b15ee07b j:\os2\xcopy.exe
```

MD5 File format

- Or a file with the checksum and filename

B2ECCC321616CB5949932C60C29990B9 *browse.exe

MD5 Checksum (may sometimes include hyphens)

i.e. B2EC-CC32-1616-CB59-4993-2C60-C299-90B9

Filename
excluding drive
and root

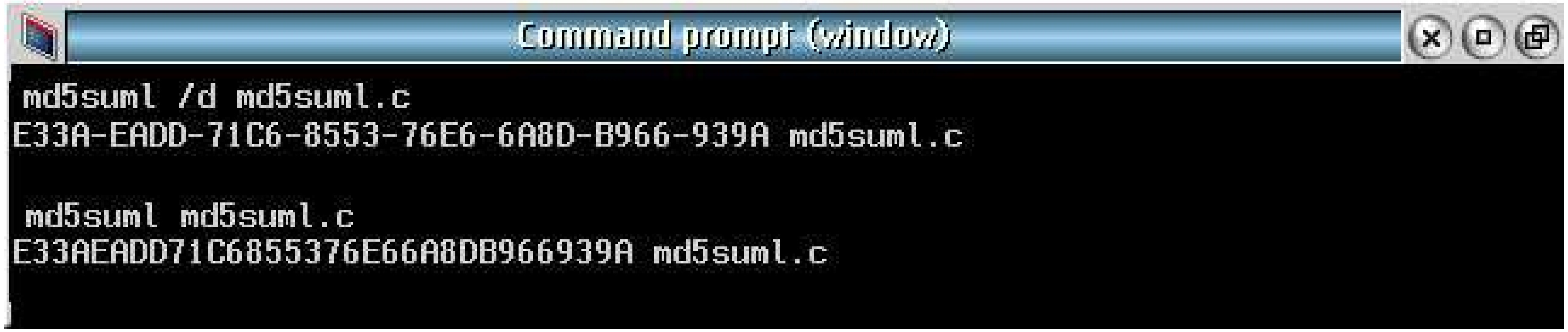
Programs using MD5 (Command line)

- Diffutils A difference utility
  - cmp.exe
  - diff.exe
  - diff1, exe
  - diff3.exe
- Xcomp. a recursive file compare utility
- There are probably more

Programs to calculate MD5 (Command line)

- MD5_OS2 (hobbes)
  - Syntax MD5 stdin/filename

- MD5SUML (hobbes) for files > 2GB
  - Syntax md5suml [/d] filename
    - -d to provide divider (hyphen) in result

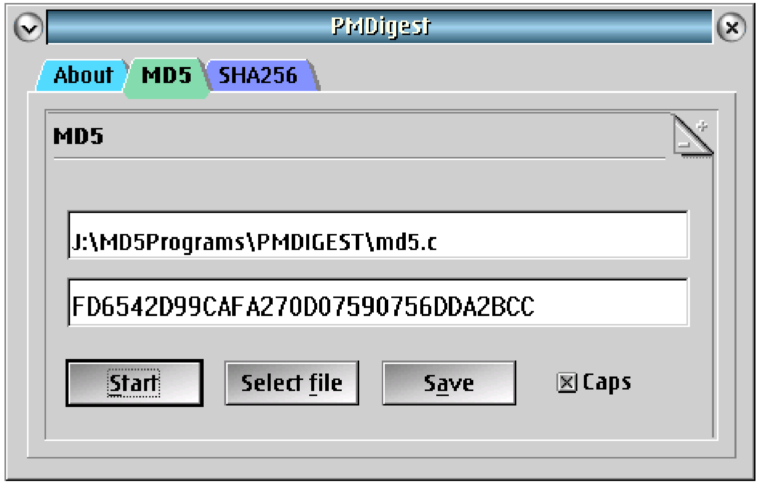


```
md5suml /d md5suml.c
E33A-EADD-71C6-8553-76E6-6A8D-B966-939A md5suml.c

md5suml md5suml.c
E33AEADD71C6855376E66A8DB966939A md5suml.c
```

Programs to calculate MD5 (PM)

PMDIGEST



J:\MD5Programs\PMDIGEST\md5.c

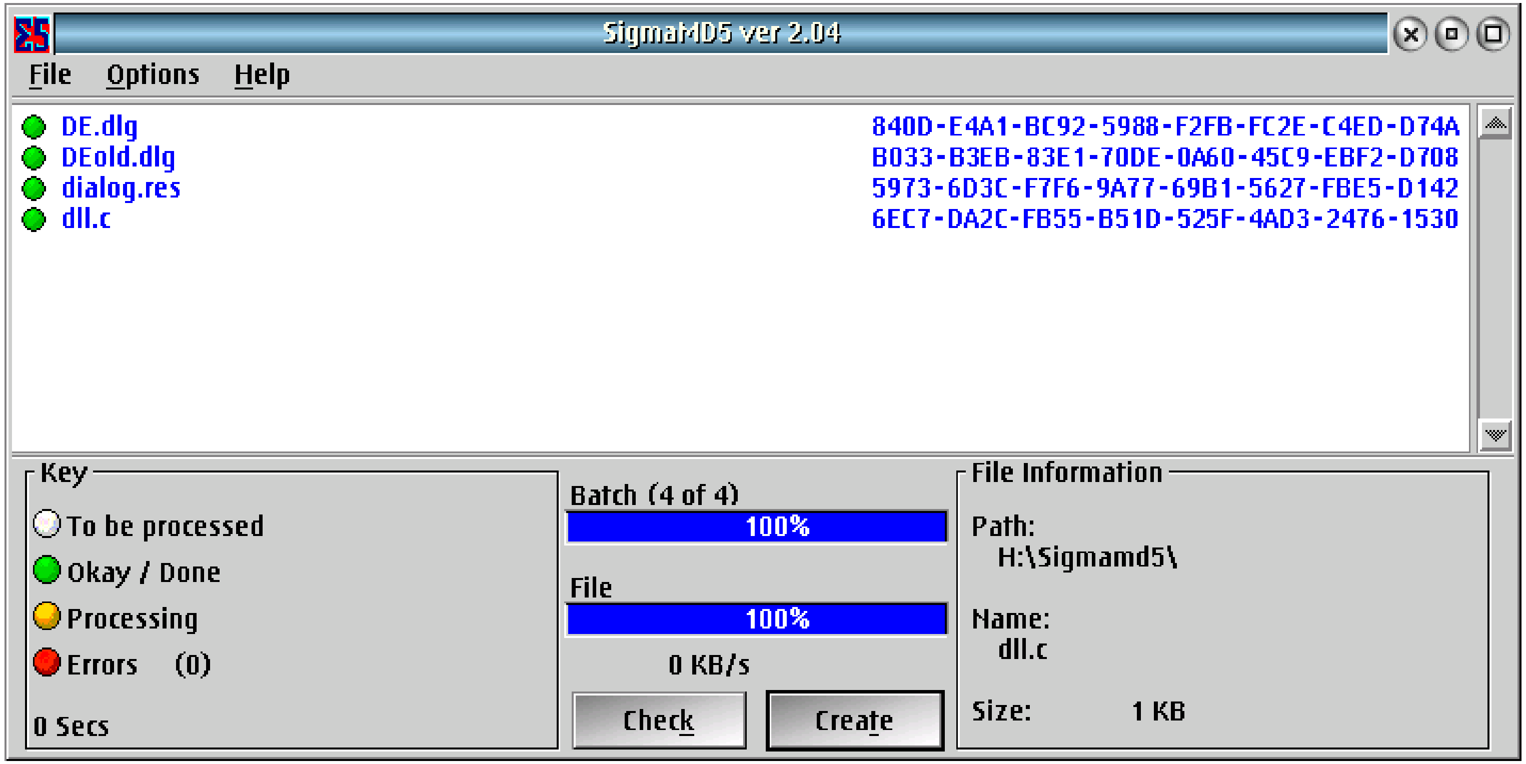
FD6542D99CAFA270D07590756DDA2BCC

Program can generate a MD5 or a SHA256 checksum for a single file.

Optionally Hex characters as upper case

Programs to calculate MD5 (PM)

SigmaMD5



```
● DE.dlg       840D-E4A1-BC92-5988-F2FB-FC2E-C4ED-D74A
● DEold.dlg    B033-B3EB-83E1-70DE-0A60-45C9-EBF2-D708
● dialog.res   5973-6D3C-F7F6-9A77-69B1-5627-FBE5-D142
● dll.c        6EC7-DA2C-FB55-B51D-525F-4AD3-2476-1530
```

Key

○ To be processed
● Okay / Done
● Processing
● Errors (0)

0 Secs

Batch (4 of 4) 100%

File 100%

0 KB/s

File Information

Path: H:\Sigmamd5\

Name: dll.c

Size: 1 KB

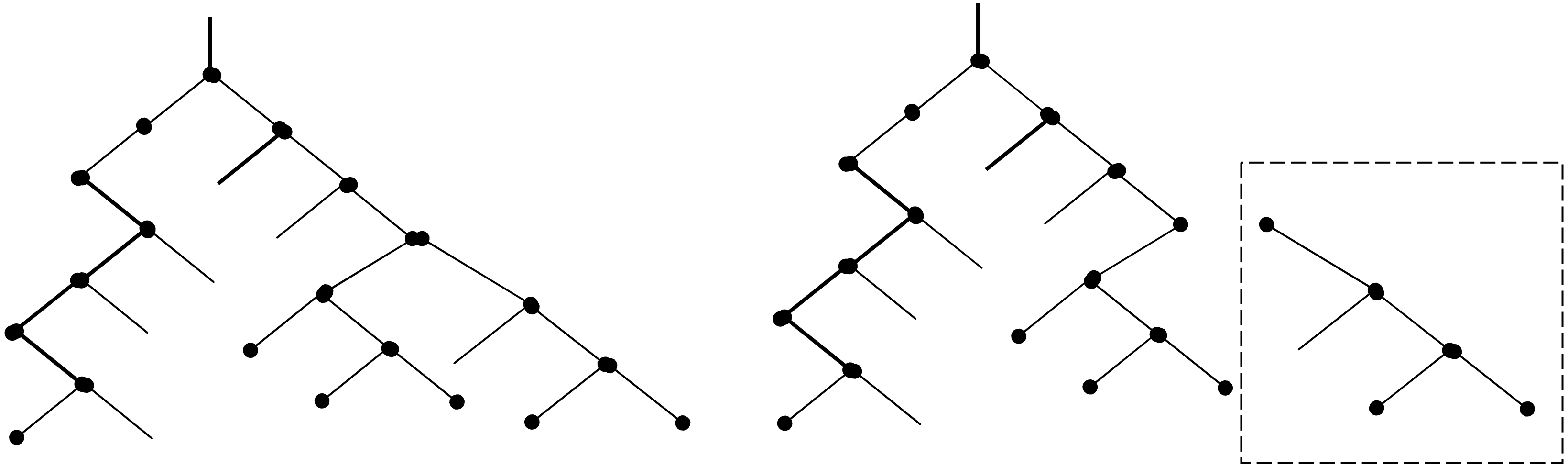
Programs to calculate MD5 (PM)

SigmaMD5:

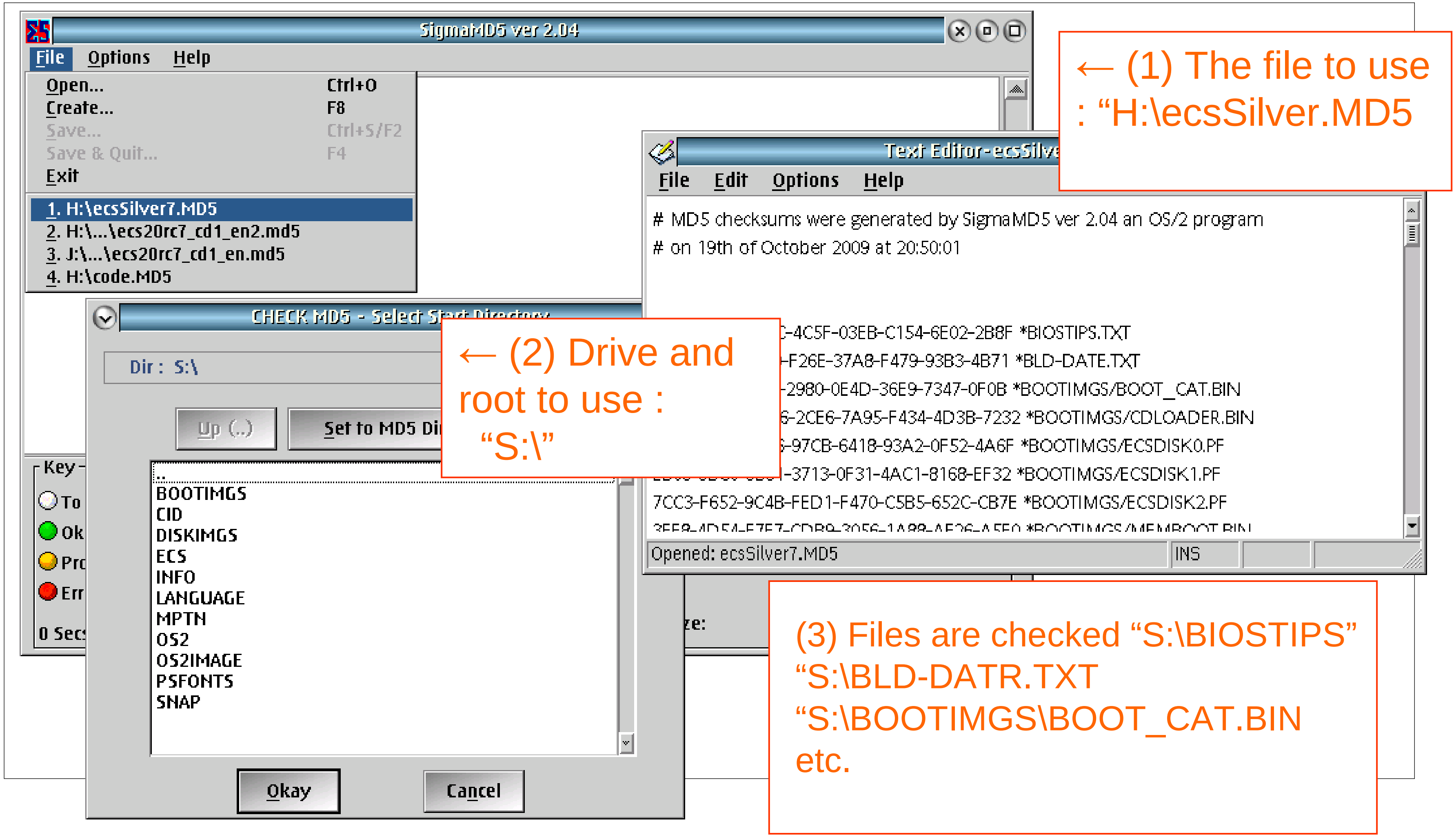
- Can calculate MD5 checksums for one or more files and save them to a MD5 list file.
- Can check the integrity of files in a MD5 list file against the MD5 checksums specified in the same file.
- Is on eCS disk 1 in \ECS\BIN\SIGMAMD5.EXE

The root

- When a MD5 file list is generated it only has a part of the path. There is NO drive letter given. So to check using the file list the starting path must be specified.



The root – an example checking eCS silver CD



```
# MD5 checksums were generated by SigmaMD5 ver 2.04 an OS/2 program
# on 19th of October 2009 at 20:50:01


C-4C5F-03EB-C154-6E02-2B8F *BIOSTIPS.TXT
-F26E-37A8-F479-93B3-4B71 *BLD-DATE.TXT
-2980-0E4D-36E9-7347-0F0B *BOOTIMGS/BOOT_CAT.BIN
6-2CE6-7A95-F434-4D3B-7232 *BOOTIMGS/CDLOADER.BIN
6-97CB-6418-93A2-0F52-4A6F *BOOTIMGS/ECSDISK0.PF
1-3713-0F31-4AC1-8168-EF32 *BOOTIMGS/ECSDISK1.PF
7CC3-F652-9C4B-FED1-F470-C5B5-652C-CB7E *BOOTIMGS/ECSDISK2.PF
3FF8-4D54-F7F7-CDB9-3056-1A99-AF26-A5E0 *BOOTIMGS/MEMBOOT.BIN
```

Opened: ecsSilver7.MD5

← (1) The file to use : “H:\ecsSilver.MD5

← (2) Drive and root to use : “S:\”

(3) Files are checked “S:\BIOSTIPS” “S:\BLD-DATR.TXT “S:\BOOTIMGS\BOOT_CAT.BIN etc.

Practical Examples

Thank You