NAME

hfind − Lookup a hash value in a hash database

SYNOPSIS

hfind [-i *db_type* **] [-f** *lookup_file* **] [-eq]** *db_file [hashes]*

DESCRIPTION

hfind looks up hash values in a database using a binary search algorithm. This allows one to easily create a hash database and identify if a file is known or not. It works with the NIST National Software Reference Library (NSRL) and the output of 'md5sum'.

Before the database can be used by 'hfind', an index file must be created with the '−i' option.

This tool is needed for efficiency. Most text-based databases do not have fixed length entries and are sometimes not sorted. The hfind tool will create an index file that is sorted and has fixed-length entries. This allows for fast lookups using a binary search algorithm instead of a linear search such as 'grep'.

ARGUMENTS

-i db_type
: Create an index file for the database. This step must be done before a lookup can be performed. The 'db_type' argument specifies the database type (i.e. nsrl-md5 or md5sum). See section below.

-f lookup_file
: Specify the location of a file that contains one hash value per line. These hashes will be looked up in the database.

-e Extended mode. Additional information besides just the name is printed. (Does not apply for all hash database types).

-q Quick mode. Instead of displaying the corresponding information with the hash, just display 0 if the hash was not found and 1 if it was. If this flag is used, then only one hash can be given at a time.

-V Display version

db_file The location of the hash database file.

[hashes]
: The hashes to lookup. If they are not supplied on the command line, STDIN is used. If index files exist for both SHA-1 and MD5 hashes, then both types of hashes can be given at runtime.

INDEX FILE

hfind uses an index file to perform a binary search for a hash value. This is much faster than using 'grep', which will do a linear search. Before a hash database is used, a corresponding index file must be created. This is done with the '−i' option to hfind.

The resulting index file will be named based on the database file name. The name will have the original name following by the hash type (sha1 or md5) followed by '.idx'. For example, creating an MD5 hash index of the NIST NSRL results in 'NSRLFile.txt-md5.idx' and the SHA-1 index results in 'NSRLFile.txt-sha1.idx'.

The file has two columns. Each entry is sorted by the first column, which is the hash value. The second column has the byte offset of the corresponding entry in the original file. So, when a hash is found in the index, the offset is recorded and then 'hfind' seeks to the entry in the original database.

The following input types are valid. For NSRL, 'nsrl-md5' and 'nsrl-sha1' can be used. The difference is which hash value the index is sorted by. The 'md5sum' value can also be used to sort and index "home made" databases. 'hfind' can take data in both common formats:

MD5 (test.txt) = 76b1f4de1522c20b67acc132937cf82e

and

76b1f4de1522c20b67acc132937cf82e test.txt

EXAMPLES

To create an MD5 index file for NIST NSRL:

hfind −i nsrl-md5 /usr/local/hash/nsrl/NSRLFile.txt

To lookup a value in the NSRL:

hfind /usr/local/hash/nsrl/NSRLFile.txt 76b1f4de1522c20b67acc132937cf82e

76b1f4de1522c20b67acc132937cf82e Hash Not Found

You can even do both SHA-1 and MD5 if you want:

hfind −i nsrl-sha1 /usr/local/hash/nsrl/NSRLFile.txt

hfind /usr/local/hash/nsrl/NSRLFile.txt
76b1f4de1522c20b67acc132937cf82e
80001A80B3F1B80076B297CEE8805AAA04E1B5BA

76b1f4de1522c20b67acc132937cf82e Hash Not Found

80001A80B3F1B80076B297CEE8805AAA04E1B5BA thrdcore.cpp

To make a database of critical binaries of a trusted system, use ’md5sum’:

md5sum /bin/* /sbin/* /usr/bin/* /usr/bin/* /usr/local/bin/* /usr/local/sbin/* > system.md5

hfind −i md5sum system.md5

To look entries up, the following will work:

hfind system.md5 76b1f4de1522c20b67acc132937cf82e

76b1f4de1522c20b67acc132937cf82e Hash Not Found

or

md5sum −q /bin/* | hfind system.md5

928682269cd3edb1acdf9a7f7e606ff2 /bin/bash

<...>

or

md5sum −q /bin/* > bin.md5

hfind −f bin.md5 system.md5

928682269cd3edb1acdf9a7f7e606ff2 /bin/bash

<...>

SEE ALSO

sorter(1)

The NIST National Software Reference Library (NSRL) can be found at www.nsrl.nist.gov.

LICENSE

AUTHOR

Brian Carrier <carrier at sleuthkit dot org>

Send documentation updates to <doc-updates at sleuthkit dot org>