Hashes

- Like a person's fingerprint
- Uniquely identifies the file based on contents
- You can't create the file from the hash
- Primary hash value used is Secure Hash Algorithm (SHA-1) specified in FIPS 180-1, a 160-bit hashing algorithm
  - 10^{45} combinations of 160-bit values
- "Computationally infeasible" to find two different files less than 2^{64} bits in size producing the same SHA-1
  - 2^{64} bits is one million terabytes

Hashes

- SHA-1 values can be cross-referenced by other products that depend on different hash values
- Other standard hash values computed for each file include Message Digest 5 (MD5), and a 32-bit Cyclical Redundancy Checksum (CRC32), which are useful in CF tools and to users outside LE

Hash Examples

Filename	Bytes	SHA-1
NT4\ALPHA\notepad.exe	68368	F1F284D5D757039DEC1C44A05AC148B9D204E467
NT4\I386\notepad.exe	45328	3C4E15A29014358C61548A981A4AC8573167BE37
NT4\MIPS\notepad.exe	66832	33309956E4DBBA665E86962308FE5E1378998E69
NT4\PPC\notepad.exe	68880	47BB7AF0E4DD565ED75DEB492D8C17B1BFD3FB23
WINNT31.WKS\I386\notepad.exe	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67
WINNT31.SRV\I386\notepad.exe	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67

Related History

- CRC concept dates from 1960's
- MD5 algorithm published in 1991
- Tripwire open source tool 1992
- Unix command "md5sum" available
- FIPS 180-1 (SHA-1) published in 1995
- Unix command "sha1sum" available
- Known File Filter project 1998
- FIPS 180-2 (SHA-512) published in 2002

SHA-1 Mathematics

- Bit sequence is padded to a multiple of 512
- Messages of 16 32-bit words, n*512, n>0
- 80 logic functions are defined that accept 3 32-bit words and produce 1 32-bit word
- 80 constants defined, 5 32-bit buffers initialized
- 80 step loop:
  - Manipulate message into 80 32-bit words
  - Use shifts, functions, addition on buffers
- 160-bit SHA is string in the 5 32-bit buffers