# Data integrity.

Article URL
Data integrity.

Author
SecurityHome.eu

Published: 12 September 2006

.

Â

If you transfer files over the internet it is possible, (espesially with large files) that there are errors in the file. Therefor you need to check the data integrity. You can do this by a checksom.

A Checksum is a computed value which depends on the contents of a block of data and which is transmitted or stored along with the data in order to detect corruption of the data. The receiving system recomputes the checksum based upon the received data and compares this value with the one sent with the data. If the two values are the same, the receiver has some confidence that the data was received correctly.

There are several technics that can be used, the most popular are:

* MD5

* CRC32

MD5

MD5 was developed by Professor Ronald L. Rivest of MIT. What it does, to quote the executive summary of rfc1321, is:

[The MD5 algorithm] takes as input a message of arbitrary length and produces as output a 128-bit "fingerprint" or "message digest" of the input. It is conjectured that it is computationally infeasible to produce two messages having the same message digest, or to produce any message having a given prespecified target message digest. The MD5 algorithm is intended for digital signature applications, where a large file must be "compressed" in a secure manner before being encrypted with a private (secret) key under a public-key cryptosystem such as RSA.

MD5 is currently a standard, Internet Engineering Task Force (IETF) Request for Comments (RFC) 1321. According to the standard, it is "computationally infeasible" that any two messages that have been input to the MD5 algorithm could have as the output the same message digest, or that a false message could be created through apprehension of the message digest. MD5 is the third message digest algorithm created by Rivest. All three (the others are MD2 and MD4) have similar structures, but MD2 was optimized for 8-bit machines, in comparison with the two later formulas, which are optimized for 32-bit machines. The MD5 algorithm is an extension of MD4, which the critical review found to be fast, but possibly not absolutely secure. In comparison, MD5 is not quite as fast as the MD4 algorithm, but offers much more assurance of data security.

Some site (like packetstormsecurity.com) have md5 checksum for there papers and programs. You can use any of the programs below to check if these docs/programs haven't been changed.

Note : Most of the times you just get the checksum (like : *c312b8c2255cb778d2fe1daf5a19593a*

) Programs like Easy MD5 creator need it in a file.
The easiest way is the UNIX way. Make a file with extention .MD5. And enter for every file :
MD5(filename.ext)=c312b8c2255cb778d2fe1daf5a19593a
That should do it.

Programs to make/check MD5:

* md5sum Linux (I know it comes with RedHat 7.1)

* Advanced CheckSum Verifier (ACSV) Windows

*CRC 32*

cyclic redundancy checksum (CRC-32)

CRC is a "digital fingerprint" of a file. With CRC32 you can "melt down" a huge 20 MB (or even much bigger) file to have a small, handy reference to it, a single 32-bit number like 7d9c42fb (hexadecimal notation) which would unambiguously reflect the entire contents of this huge file. Now if some changes to this file happened, no matter how small, maybe only a single wrong bit somewhere in the middle, a new CRC-32 calculation would yield a completely different reference number (say 3faa83bd). So there'd be no doubt about it - this is not the same file anymore. On the other hand if the reference number was still the same (7d9c42fb) you might be sure that the file hasn't changed.

If you download a folder and there is file with extention *.SFV*

you can check if the files are downloaded corectly with the programs below:

Programs to make/check crc32:
* FSUM

* Advanced CheckSum Verifier (ACSV) Windows