Netlib News: MD5 checksums

Eric Grosse

15 June 1998

We resume the Netlib News column with a series on computer network security and how it relates to the netlib mathematical software repository.

This column introduces the digital signatures that were added to netlib a couple years ago, and what protections they provide. Security hasn't been a pressing issue, because fortunately we have not been subjected to any fraud or pranks. However, security alerts are becoming more frequent in other areas of computing, and it seems prudent to put the infrastructure in place, ready to invoke when the bandits come to our territory. Next month's column will go into more detail about getting and installing PGP, the Pretty Good Privacy cryptographic software that we're using. Future columns will describe how authors can automatically update their software in the collection, by sending PGP-signed email to the server.

Each directory in netlib contains a file `MD5`, which is a simple text file containing a list of filenames and checksums. You can compute for yourself these "message digest 5" checksums, also known as "one-way hash functions", using `/netlib/crc/md5sum.c`. It is astronomically difficult to create a file with predetermined MD5 checksum. In contrast, the 32-bit CRC cyclic redundancy checksum, which is faster to compute and therefore preferred in situations where deliberate tampering is not a threat, can easily be forged. Indeed, the software that creates `/netlib/bibnet` bibliography files puts a CRC in the file, tweaking things so that the final file contains its own checksum!

Go ahead and try this out. Download `md5sum.c`, compile, and compare with /netlib/crc/MD5, confirming that you've correctly installed the program. (Please let me know if you have any difficulty.) The slightest change to a file, even from the one-byte Unix line terminators to the two-byte Windows line terminators, will result in a different checksum, so you do need to take care to do the confirmation on exact copies before converting to local system formats.

Already you have a useful tool. Print the checksum of some file such as a proof or invention, give that to a trusted party for safekeeping, and you can later prove to any reasonable person's satisfaction that you really did have the original file at the time you handed off the checksum. Note that the trusted party needn't ever see the file, or save anything more than the small checksum.

A clever bandit that maliciously changes a netlib file, either in transit or on a mirror site, could simultaneously change the MD5 file. Next time, we'll look at the digital signature at the bottom of netlib's MD5 files, which let's you protect yourself against such threats.

Recent additions to netlib

SuperLU_MT in `scalapack/prototype` solves a sparse linear system by Gaussian elimination with partial pivoting, exploiting multithreading if available. Currently, the LU factorization is parallelized on shared memory machines with POSIX threads or Sun Ultra Enterprise servers, DEC Alpha Servers, SGI Power Challenge, SGI/Cray Origin2000, Cray C90/J90.

`linalg/amd/` is the approximate minimum degree ordering of a sparse symmetric matrix by Davis Amestoy Duff Reid.

`cephes/c9x-complex.shar` by Moshier contains a runtime library needed to support type complex in the C9X standard.

Modern computers have deeply pipelined functional units and two or more layers of memory hierarchy. The `atlas` project (Whaley Dongarra) has produced a (BLAS3) `gemm` that automatically tunes itself for such architectures.

Frigo and Johnson continue to expand the scope and languages of `transform/fftw.tar.gz`, an FFT in one or more dimensions using C, Cilk, and MPI. Casanova issued a new release of `netsolve`, a system for client/server numerical computing. The sqrt monotonicity test in the popular floating point test `paranoia` was patched recently; a negation had been lost in the translation from Basic to Pascal and C. David Gay's popular programs `ampl, f2c, fp` are constantly undergoing small improvements.

The `access` library contains tools for getting files from netlib when your favorite web browser is not up to the task. The `unshar.c` was improved to cope with additional shar variants; a command `webget.c` was added to work around bugs in existing browsers regarding transfer of compressed files.

Journal algorithms:
numeralgo/na13 Lucet: linear-time Legendre transform
toms/771 Brankin Gladwell, rksuite_90: ODE
toms/772 Renka, stripack: Delaunay triangulation on a sphere
toms/773 Renka, ssrfpack: surface under tension on a sphere
toms/774 Facchinei Judice Soares: box-constrained optimization

Some PGP fingerprints associated with netlib:
`netlib-bl 28 AA 7C 79 00 AB 2F 58 4A 77 BA 04 5B 2F C3 C2`
`ehg 28 A9 32 6F 02 2E 7A F9 C4 6B 26 B5 21 5C 16 28`
See my home page or the MIT public keyservers for the public keys themselves.

Eric Grosse is at the Computing Sciences Research Center, Bell Labs, Lucent Technologies, Murray Hill NJ 07974, USA. `http://cm.bell-labs.com/who/ehg/`