----

ooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooX

Message 64 (1594 chars)
Return-Path: <hector@DB.Stanford.EDU>
Received: from Coke.Stanford.EDU by EE.Stanford.EDU with SMTP (5.67b8/25-eef) id AA04126; Fri, 27 Jan 19   ...(
Received: by Coke.Stanford.EDU (5.57/25-DB-eef) id AA27352; Fri, 27 Jan 95 12:06:21 -0800
Message-Id: <9501272006.AA27352@Coke.Stanford.EDU>
To: shankle@ee.Stanford.EDU
Subject: qual question
Date: Fri, 27 Jan 95 12:06:20 -0800
From: hector@DB.Stanford.EDU
X-Mts: smtp

EE Qual Question   c S
Hector Garcia-Molina, 1995

You wish to compare two files stored on separate computers, to see if they are
identical.

PART I: How could you compare the files?

POSSIBLE ANSWER:
(1) Read file on one machine, send all bytes across network, do a
    byte by byte compare
(2) Compute a "check sum" or signature at one machine, send it across network,
    and compare to signature of other file.
(3) Assuming files were identical initially, keep a log of changes.
    Then compare logs.
(4) Compress the file before sending across network.

PART II: For strategy (2) above, compute the probability that the
    test "fails". Assume each file has p bits, signature is s bits.

POSSIBLE ANSWER:
Consider a given file X with signature Sx being compared against a file Y.
There are $2^{**}p$ possible Y files. Of these,
$2^{**}p/2^{**}s$ have the same signature as Sx. Of these, $(2^{**}p/2^{**}s) - 1$
will cause our test to fail. (The remaining one with Sx signature
is identical to X so the test does not fail.) Thus, the probability
that our test gives an incorrect answer is $[(2^{**}p/2^{**}s) - 1]/(2^{**}p)$.
If p is substantially larger than s, this simplifies to $1/(2^{**}s)$.