X-Sieve: CMU Sieve 2.2
Date: Mon, 23 Apr 2007 10:33:14 +1000
From: Scott Contini <scontini@ics.mq.edu.au>
User-Agent: Thunderbird 1.5.0.8 (Windows/20061025)
To: hash-function@nist.gov
Subject: Hash Algorithm Requirements and Evaluation Criteria
X-Proofpoint-Virus-Version: vendor=fsecure engine=4.65.5502:2.3.11,1.2.37,4.0.164
definitions=2007-04-23_01:2007-04-19,2007-04-22,2007-04-23 signatures=0
X-PP-SpamDetails: rule=spampolicy2_notspam policy=spampolicy2 score=0 spamscore=0
ipscore=0 phishscore=0 adultscore=0 classifier=spam adjust=0 reason=mlx engine=3.1.0-
0703060001 definitions=main-0704220047
X-PP-SpamScore: 0
X-NIST-MailScanner: Found to be clean
X-NIST-MailScanner-From: scontini@ics.mq.edu.au

Dear National Institute of Standards and Technology,

I have some concerns about the draft requirements and evaluation criteria for a new hash standard. Briefly, my concerns are the following:

1. The security goal of the outputs being indistinguishable from a random oracle is not well defined.
2. NIST should leave the door open for provable (i.e. having a security reduction) designs to be considered.

I elaborate on both of these concerns below.

In section C.1, NIST currently states that one of the security measurements of candidate algorithms is

> "The extent to which the algorithm output is indistinguishable from a random oracle."

I'm not sure how one can make any conclusions about how outputs are distinguishable from random oracles if one does not also consider how the inputs are chosen (if I am wrong on this, please give more specific details on what is meant). Furthermore, there needs to be a very precise, mathematical description of what it means to successfully distinguish from a random oracle: otherwise judging which hash functions are behaving best becomes completely subjective. A fair competition should have rules than can be objectively evaluated, which requires more precise definitions that so far have not been provided.

In regard to provable designs, although section B.1 does mention the possibility of a security reduction proof, the bulk of the document seems to suggest that the goal of the hash function is to achieve multiple properties including random oracle behavior. Historically, provable hash properties and random oracle behavior have been disjoint topics. Especially taking into consideration research initiated by Canetti, Goldreich, and Halevi that shows a separation between random oracle security and security in the real world, in the last ten years there has been a migration away from random oracle assumptions by a large portion of the research community. The trend has been towards developing protocols where the security requirement of the hash function is one specific and achievable property, such as collision resistance (Examples include the Cramer-Shoup signature and encryption algorithms). This more theoretically sound approach to hashing suggests that different hash functions should be developed to achieve different

properties rather than a "one hash to solve all problems" approach, like the way hashing is done today. I therefore request, in order to accommodate emerging research, that your competition has multiple categories for hash submissions according to what the submitted hash functions are intended to achieve. Examples of such categories include:

- Provable hash functions aimed at providing collision resistance only.
- Provable hash functions aimed at providing preimage resistance only.
- Provable hash functions aimed at behaving like a PRF when keyed through the IV (or keyed some other way).
- Heursitic hash functions that are aimed at random oracle emulation.

The last one, heuristic random oracle emulation, is really what designs like SHA-1 have been intending to achieve. Although nobody has precisely defined what this means, it is implicitly assumed that that would imply preimage resistance, second preimage resistance, and collision resistance. However, the goal has problems from a theoretical point of view. For example, how can one really define what random oracle emulation means and show that it implies collision resistance when we really cannot even formally define collision resistance in the complexity theory model according to the way we are doing hashing today in practice? Thus it seems that any such definition must appeal to "human ignorance" (see Phil Rogaway's Vietcrypt 2006 paper). In short, the theoretical problems with the way we are doing hashing today motivates a competition which allows more theoretically sound hash solutions to be considered. I'm quite sure that there are organizations who would prefer the more theoretically sound approach, even if it comes at relatively small speed sacrifices.

Thank you for taking the time to consider public feedback for your draft standard.

Sincerely,

Scott Contini
Department of Computing
Macquarie University