Workshop Report
The Second Cryptographic Hash Workshop

University of California, Santa Barbara
August 24-25, 2006

Report prepared by
James Nechvatal and Shu-jen Chang

Information Technology Laboratory,
National Institute of Standards and Technology,
Gaithersburg, MD 20899

Available online:
http://csrc.nist.gov/groups/ST/hash/documents/SecondHashWshop_2006_Report.pdf

1. Introduction

The Second Cryptographic Hash Workshop was held on Aug. 24-25, 2006, at University of California, Santa Barbara, in conjunction with Crypto 2006. 210 members of the global cryptographic community attended the workshop. The workshop was organized to encourage hash function research and discuss hash function development strategy.

2. Workshop Program

The workshop program consisted of a day and a half of presentations of papers that were submitted to the workshop and panel discussion sessions. The main topics of the workshop included survey of hash function applications, new structures and designs of hash functions, cryptanalysis and attack tools, and development strategy of new hash functions. The program is available at http://csrc.nist.gov/groups/ST/hash/second_workshop.html .

This report only briefly summarizes the presentations, because the above web site for the program includes links to the speakers' slides and papers. The main ideas of the discussion sessions, however, are described in considerable detail; in fact, the panelists and attendees are often paraphrased very closely, to minimize misinterpretation. Their statements are not necessarily presented in the order that they occurred, in order to organize them according to NIST's questions for each session.

3. Session Summary

Session 1: Papers - New Structures of Hash Functions

Session 1 included three talks dealing with new structures of hash functions. Orr Dunkelman of the Computer Science Department of Technion - Israel Institute of Technology began the session with a talk on a framework for iterative hash functions. He stated that the Merkle-Damgard construction, per se, is deficient. He proposed to add two parameters, the number of bits hashed so far and a salt. He noted that the number of bits hashed so far is a parameter which most hash functions keep anyway, so the salt is the only real addition. The added parameters enhance security against collision and pre-image attacks. In particular, inclusion of a salt introduces randomization, making it more difficult to find two messages with the same hash value.

Shoichi Hirose of the University of Fukui proposed a construction for double block length hash functions. With his approach, hash functions are constructed from block ciphers such as the AES. A block cipher is used to construct a compression function. The output length of the hash function is twice the output length of the block cipher. The key length of the block cipher must exceed block length; thus, e.g., AES with 192-bit or 256-bit keys can be used. It is also possible to use a smaller compression function in place of the block cipher.

Thomas Ristenpart of the University of California, San Diego Security and Cryptography Laboratory proposed replacing the Merkle-Damgard transform with a multi-property-preserving domain extension transform. The goal is to build hash functions to be secure for as many applications as possible. He gives an example of a multi-property-preserving transform, called Enveloped Merkle-Damgard (EMD). This is shown to have some provable security properties.

Session 2: Papers - Hash Functions in Practice

Session 2 included two talks dealing with hash functions in practice. Shin'ichiro Matsuo of NTT DATA Corporation began the session by discussing current usage of hash functions and proposing a classification of hash functions. In particular, he proposed that the concept of collision resistance be refined by introducing a qualitative index and a quantitative index. The new classification scheme is thus a two-dimensional matrix. Usages of hash functions can be mapped into this table.

Shai Halevi of IBM T.J. Watson Research Center discussed randomized hashing. This involves choosing a random salt and then hashing the message and the salt together to get the hash value. A fresh salt is chosen for each signature. This provides resistance against off-line collision attacks. It also provides a safety net, in the event that the underlying hash function exhibits weaknesses that were not apparent when it was adopted.

Session 3: Panel/Open Discussion - SHA-256 Today and Maybe Something Else in a Few Years: Effects on Research and Design

This panel was chaired by Paul Hoffman of VPN Consortium and Arjen Lenstra of Ecole Polytechnique de Lausanne IC LACAL. The panelists included Niels Ferguson from Microsoft, Antoine Joux from the Universite de Versailles-Saint-Quentin-en-Yvelines, Bart Preneel from Katholieke Universiteit Leuven, Ron Rivest from the Massachusetts Institute of Technology, and Adi Shamir from the Weizmann Institute of Science.

This panel and the audience addressed the following issues related to SHA-256 and future hash functions:

1. What problems with SHA-1 and SHA-256 do we really face?

Shamir stated that the main problem with SHA-1 is one of perception; the cryptographic problem is not as serious as perceived by some members of the public.

Preneel noted that MD5 has been known to be insecure for a considerable time period but is still in use; SHA-1 is analogous. He later noted that since the design principles of SHA-256 have not been made public, it is difficult to determine how much security SHA-256 provides beyond that provided by SHA-1.

Rivest suggested that future designs should incorporate defenses against attacks going beyond those of the SHA family. In particular, round number should be parameterized.

Joux summarized the preceding by concluding that there is considerable doubt remaining as to how future hash functions should be designed, and precisely what properties should be sought in them.

2. What properties of hash functions do we know we need for the long term?

Ferguson noted that hash functions must be resilient, since legacy systems continue to use hash functions even after weaknesses have appeared. He also said that hash functions are used as random mappings, so this should be a design requirement.

Rivest proposed that two processing modes be considered for future hash functions: the usual one-pass stream mode, and a second in-memory mode that would permit multiple passes over the data or processing of the message as a whole.

Preneel emphasized that a hash function must give resistance to collisions.

Joux questioned whether the notion of collision resistance could be formally specified for fixed hash functions.

Shamir stated that collision resistance should hold even when outputs are truncated, e.g., from 256 to 128 bits.

3. Should we develop one all-purpose hash or several special-purpose hash functions?

Several panelists discussed the issue of one versus several hash functions. It was pointed out that each function requires implementation, testing, and certification, greatly expanding cost if multiple functions are approved. The general consensus seemed to be that one function is optimal.

Preneel agreed with the one-function consensus, but wanted variants produced by using the one function with multiple modes and numbers of rounds.

Shamir agreed with the one-function, multiple-mode paradigm, but stated that modes should not alter the internal structure of a compression function.

Joux stated a preference for a single function that emulates a random oracle. However, he believes that streaming modes alone will not suffice in this regard, and that an in-memory variant will be needed.

Ferguson argued that variants using multiple passes or randomly accessing the whole message will be difficult to implement from an engineering point of view.

Rivest stated that the algorithm chosen should be available in several different strengths, by varying the round number. He also suggested that reduced-round versions be included in the algorithm specification, so as to facilitate analysis.

The topic of low-end versus high-end variants was also discussed. The consensus seemed to be that NIST should concentrate on the high end of the computational spectrum. Users employing low-end implementations should do so at their own risk.

4. How do we design the next algorithm(s)?

The question arose as to whether completely new algorithms are needed, or whether current algorithms and design paradigms merely need modification.

Shamir stated that the current set of algorithms is seriously flawed. However, he also believes that there is no alternative set of algorithms ensuring more security than existing algorithms. He concluded that modifications of existing algorithms should be given serious consideration.

Ferguson drew upon the AES competition as a source of enlightenment. He believes that the state of block cipher design was enhanced on a feedback basis by the AES competition. He stated that a similar phenomenon would characterize a hash function competition, i.e., our knowledge of hash function design would improve during the competition.

Shamir amplified on his previous remarks by noting that a hash competition would entail two kinds of proposals: modifications of existing algorithms and paradigms, and proposals of a new nature. He asked how it would be possible to evaluate the proposals based on new ideas, and compare the results to proposals based on more traditional approaches.

Rivest replied to the previous point by noting that although it might be unlikely that an entirely new algorithm would be selected, creation of such algorithms advances the state of the art of design.

The question was raised as to whether entries to a competition could be fruitfully revised based on other entrants' ideas. In theory this might allow for a combination of the best ideas. However, skepticism was expressed as to whether this would actually occur during a competition.

Joux noted that a competition would encourage further examination of SHA-256.

The subject of performance versus security was raised. It was noted that there is a trade-off between performance and security, and that a fixed round number encourages entrants to choose a close-to-the-bone round number, in order to maximize performance. Shamir and Rivest noted that parameterizing the number of rounds can solve this problem.

A workshop participant asked what levels of security should be required, and whether SHA-512 would be needed to complement AES-256. Ferguson thought SHA-256 would go well with AES-256. Shamir expressed the opinion that to use AES-256 is overkill unless one is seriously worried about quantum computers. He thought the combination of AES-128 with SHA-256 is adequate for any reasonable security applications.

The panelists returned to the question of modification of existing algorithms/paradigms versus new ideas. The consensus seemed to be that the time frame for choosing an algorithm made going with a completely new idea problematical. Preneel expressed the opinion that it might be appropriate to take some risks in examining alternatives to the Merkle-Damgard paradigm.

Keynote Speech: Message Modification, Neutral Bits and Boomerangs: From Which Round Should we Start Counting in SHA?

Antoine Joux of University of Versailles, Saint-Quentin-en-Yvelines delivered the keynote speech for the workshop. He began by discussing earlier cryptanalytic work on SHA-0, dating to 1998. In this context he treated the more general question of how to count complexity in differential cryptanalysis. He reviewed basic attacks on SHA-0, and then explained how SHA-0 metamorphosed into SHA-1, with an attendant increase in security against differential attacks. He explained how to find differential collisions in SHA-0, and why this is more difficult in SHA-1. He also explained creation of low-weight vectors. Later he explained boomerang attacks and mentioned automatic tools for finding non-linear characteristics. He noted that as a result of recent advances, SHA-1 at present is weaker than SHA-0 was in 1998.

Session 4: Papers - New Designs of Hash Functions

Session 4 included two talks dealing with new designs of hash functions. Gilles Van Assche of STMicroelectronics began the session by discussing an approach to design of cryptographic hash functions based on the approach underlying the PANAMA hash function. This provides an alternative to the Merkle-Damgard construction. A concrete example of the new construction is RadioGatun, termed a belt-and-mill hash function. He asserted that RadioGatun is competitive with SHA-1 in terms of performance, while correcting the security deficiencies of SHA-1.

Markku-Juhani O. Saarinen of Royal Holloway, University of London discussed LASH, a cryptographic hash function based on the Miyaguchi-Preneel construction. LASH uses modular matrix multiplication as the main component. He asserted that LASH is competitive with the SHA family in terms of performance, while correcting the security deficiencies of the SHA family.

Session 5: Papers/Panel - Cryptanalysis and Attack Tools

Session 5 consisted of paper presentations and a panel discussion on cryptanalysis and attack tools chaired by John Kelsey of NIST. Kelsey began the session with a talk on background and overview of cryptanalysis and attack tools pertaining to hash functions. He discussed general approaches to finding collisions in compression functions and whole hash functions. He also discussed finding differential paths and single-block collisions, as well as message modification and use of multiple blocks.

Christian Rechberger of the Faculty of Computer Science of Graz University of Technology discussed recent progress in searching automatically for complex characteristics for the SHA family. To illustrate the method, he finds a 2-block collision for 64-step SHA-1. He also discussed the potential for extending the method to other hash functions, including SHA-2.

Werner Schindler of Bundesamt fur Sicherheit in der Informationstechnik (BSI), Bonn (Federal Office for Information Security, Germany) discussed methodology to determine precise probabilities for differential near-collision paths in MD-5 type hash collision attacks. As an illustration, the method is fully realized for calculating the overall probabilities of three near-collision paths for MD5. He discussed the potential for extending these calculations to the case of SHA-1.

Philip Hawkes of Qualcomm Australia discussed automated search for round 1 differentials for SHA-1. His goal is to find new differential paths through the first round of SHA-1. Automation of the search process is achieved through the use of two search trees, one forward and one reverse. The differential paths in the leaves of the two trees are compared; a match is sought. As yet, comparison/matching has not been implemented.

Makoto Sugita of Cryptography Research and Evaluation Group IT Security Center discussed Wang's work on cryptanalysis of SHA-1, and noted that many details of Wang's work are not available. He addresses two areas arising in this context: determination of sufficient conditions, and message modification technique. His treatment of the latter employs Grobner basis based cryptanalysis of SHA-1.

Session 6: Papers - More New Designs of Hash Functions

Session 6 included three talks dealing with new designs of hash functions. Danilo Gligoroski of Norwegian University of Science and Technology began the session by discussing a new family of cryptographic hash functions, Edon-R. This is a class of hash functions with variable output lengths. It is based on quasigroups and quasigroup string transformations. He conjectured that finding collisions or preimages for Edon-R is equivalent to the apparently intractable problem of solving a system of equations in shapeless quasigroups.

Kristin Lauter of Microsoft Research discussed the construction of provable collision resistant cryptographic hash functions from expander graphs. In particular, she uses a family of optimal expander graphs called Ramanujan graphs. An example is the graph of supersingular elliptic curves modulo a prime. In this case, collision resistance follows from hardness of computing isogenies between supersingular elliptic curves.

Chris Peikert of Massachusetts Institute of Technology discussed a new family of provably secure, collision-resistant cryptographic hash functions. He uses Fast Fourier Transforms to achieve ideal diffusion properties, along with a random linear function to achieve confusion. He can show formally that finding collisions in the resulting family is at least as hard as approximating the shortest vector in certain lattices.

Session 7: The Way Forward

Session 7 was a wrap-up session chaired by Bill Burr, manager of the Security Technology Group, Computer Security Division of NIST. This session consisted of a short presentation, a report of new results, a discussion of a proposed timeline for developing new hash functions, a summary of the workshop, and an open discussion for the way forward.

Stuart Haber of HP Labs gave a short presentation to remind the workshop attendees of an old solution to a new problem, namely, how one can ensure long term integrity to digital objects in the face of deprecated hash functions used in creating the original digital objects.

Lisa Yin, an independent security consultant in Connecticut, presented forgery and partial key recovery attacks on HMAC and NMAC using hash collisions. She also presented some new observations on the second pre-image resistance of MD5 and reduced SHA-1. She concluded that at present, HMAC-MD4 is not secure, but there are no immediate practical threats against HMAC-MD5 or HMAC-SHA1.

Elaine Barker of NIST discussed a proposed timeline for the development of new hash functions. She stated that FIPS 180-2 will be reviewed in 2007 and again in 2012, so 2012 would be the logical time to have a new hash function standardized. The timeline was developed with this goal in mind. NIST will publish information related to the competition of new hash functions, including minimum acceptability requirements, evaluation criteria, and submission requirements. These will be subject to public comment. The competition process was described, and is expected to be completed by 2012. Presentation of the proposed timeline was followed by a discussion. Barker tried to get feedback on the question of how many more hash research workshops would be needed before NIST initiates the competition, and when would be a suitable time and what would be a suitable length for the next workshop. There was a strong sentiment that beginning the competition relatively soon would help focus the cryptographic community, and benefit the research that's needed for the competition. In addition, it was suggested that more time be devoted to the public analysis of the candidate algorithms than was originally allocated. It was also suggested that "borrowing ideas" should be considered; that is, submitters should be allowed to re-submit proposals based on feedback and the study of other proposals. Donna Dodson of NIST reminded the workshop attendees that cross pollination may be a good idea, but trying to maintain fairness to everyone and handle the Intellectual Property issue would be a challenge for NIST. She emphasized that the process that NIST has in place is all about fairness, and that process can take time, but it is a principle that NIST would adhere to, as exemplified by the success of the AES competition.

Bill Burr of NIST concluded with a wrap-up. He noted that there is a need for hash functions to be applicable to a wide variety of applications. He suggested that it might be good for proposed hash functions to parameterize the number of rounds, as well as accommodating a variety of hash output sizes. He stated that NIST plans on an AES-like competition to select a new hash function. In the evaluation process, he stated that NIST should emphasize security over performance, although performance is easier to measure. He also stated that maybe only one new hash function should be chosen, but various modes may be standardized. In addition, the possibility of allowing more extensive changes to finalist proposals than were allowed in the AES should be considered. NIST will develop a plan for moving forward based on the feedback gathered from this workshop.