Moni Naor (Ed.)

LNCS 4515

Advances in Cryptology – EUROCRYPT 2007

26th Annual International Conference on the Theory and Applications of Cryptographic Techniques
Barcelona, Spain, May 2007, Proceedings

INTERNATIONAL ASSOCIATION FOR CRYPTOLOGIC RESEARCH

iacr

Springer

Moni Naor (Ed.)

Advances in Cryptology - EUROCRYPT 2007

26th Annual International Conference on the Theory and Applications of Cryptographic Techniques
Barcelona, Spain, May 20-24, 2007
Proceedings

Springer

Volume Editor

Moni Naor
Weizmann Institute of Science
Department of Computer Science and Applied Mathematics
Rehovot 76100 ISRAEL
E-mail: moni.naor@weizmann.ac.il

Library of Congress Control Number: 2007926705

CR Subject Classification (1998): E.3, F.2.1-2, G.2.1, D.4.6, K.6.5, C.2, J.1

LNCS Sublibrary: SL 4 – Security and Cryptology

ISSN 0302-9743
ISBN-10 3-540-72539-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72539-8 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12064380 06/3180 5 4 3 2 1 0

Preface

These are the proceedings of Eurocrypt 2007, the 26th Annual IACR Eurocrypt Conference. The conference was sponsored by the International Association for Cryptologic Research (IACR; see www.iacr.org), this year in cooperation with the Research Group on Mathematics Applied to Cryptography at UPC and the Research Group on Information Security at UMA. The Eurocrypt 2007 Program Committee (PC) consisted of 24 members whose names are listed on the next page.

The PC decided on several policies: zero PC papers - no Program Committee member could submit papers; optional anonymity - authors could choose to anonymize their papers or not. Anonymous papers were treated as usual, i.e., the author's identity was not revealed to the PC. The submission software used was "Web Submission and Review Software" written and maintained by Shai Halevi. There were 173 papers submitted to the conference and the PC chose 33 of them. Each paper was assigned to at least three PC members, who either handled it themselves or assigned it to an external referee. After the reviews were submitted, the committee deliberated both online for several weeks and finally in a face-to-face meeting held in Paris. In addition to notification of the decision of the committee, authors received reviews. Our goal was to provide meaningful comments to authors of all papers (both those selected for the program and those not selected). The default for any report given to the committee was that it should be available to the authors as well.

The committee decided to give the Best Paper Award to Shien Jin Ong and Salil Vadhan for their paper "Zero Knowledge and Soundness are Symmetric." In addition the PC chose two more notable papers for invitation to the *Journal of Cryptology*. These are "Chosen-prefix Collisions for MD5 and Colliding X.509 Certificates for Different Identities," by Marc Stevens, Arjen Lenstra and Benne de Weger, and "An $L(1/3 + \varepsilon)$ Algorithm for the Discrete Logarithm Problem for Low-Degree Curves," by Andreas Enge and Pierrick Gaudry. The conference program included two invited lectures: by Jacques Stern (IACR Distinguished Lecture) titled "Cryptography from A to Z" and by Victor Miller titled "Elliptic Curves and Cryptography: Invention and Impact."

I wish to thank all the people who made the conference possible. First and foremost the authors who submitted their papers. The hard task of reading, commenting, debating and finally selecting the papers for the conference fell on the PC members. I am indebted to the committee members' collective knowledge, wisdom and effort. I have learned a lot from the experience. The committee also used external reviewers, whose names are listed on the following pages, to extend the expertise and ease the burden. My deepest gratitude to them as well. I thank Shai Halevi for handling the submissions and reviews server and Michel Abdalla for organizing the PC Meeting in Paris. I am grateful to previous PC Chairs who have shared their experiences with me. Finally, the Eurocrypt General Chairs Javier López and Germán Sáez and the local organizing committee Monica Breitman, Paz Morillo and Jorge L. Villar deserve many thanks from all the IACR community for the organization of the conference.

March 2007

Moni Naor

Eurocrypt 2007

Barcelona, Spain, May 20–24, 2007

Sponsored by the *International Association for Cryptologic Research.*

Organized in cooperation with the
Technical University of Catalonia (UPC) and the University of Malaga (UMA).

General Chairs

Javier López and Germán Sáez

Program Chair

Moni Naor, Weizmann Institute of Science

Program Committee

Michel Abdalla	ENS and CNRS, Paris
Anne Canteaut	INRIA-Rocquencourt
Dario Catalano	University of Catania
Jung Hee Cheon	Seoul National University
Stefan Dziembowski	Warsaw University and University of Rome "La Sapienza"
Serge Fehr	CWI, Amsterdam
Marc Fischlin	TU Darmstadt
Jens Groth	UCLA
Shai Halevi	IBM T.J. Watson Research Center
Yuval Ishai	Technion
Joe Kilian	Rutgers University
Anna Lysyanskaya	Brown University
Alexander May	TU Darmstadt
Steven Myers	Indiana University
Moni Naor	Weizmann Institute of Science
Phong Nguyen	ENS and CNRS, Paris
Jesper Buus Nielsen	University of Aarhus
Giuseppe Persiano	University of Salerno
Ron Rivest	MIT
Alon Rosen	Harvard
Eran Tromer	MIT

Xiaoyun Wang	Tsinghua University
Brent Waters	SRI
Stefan Wolf	ETH, Zurich

External Reviewers

Ben Adida
Roberto Araujo
Thomas Baignères
Boaz Barak
Amos Beimel
Ian F. Blake
Carlo Blundo
Alexandra Boldyreva
Xavier Boyen
Jan Camenisch
Jean Camp
Ran Canetti
Melissa Chase
Liqun Chen
Benoît Chevallier-Mames
Joo Yeon Cho
Paul Crowley
Chris Crutchfield
Ivan Damgård
Cécile Delerablée
Alex Dent
Claus Diem
Jintai Ding
Christophe Doche
Martin Döring
Orr Dunkelman
Jean-Charles Faugère
Nelly Fazio
Matthias Fitzi
Pierre-Alain Fouque
Fabien Galand
Steven Galbraith
Clemente Galdi
Pierrick Gaudry
Rosario Gennaro
Vipul Goyal
Tim Güneysu
Robbert de Haan
Iftach Haitner
Guillaume Hanrot
Danny Harnik
Alex Healy
Martin Hirt
Dennis Hofheinz
Susan Hohenberger
Thomas Holenstein
Nick Howgrave-Graham
Vasyltsov Ihor
Stanislaw Jarecki
Antoine Joux
Pascal Junod
Jonathan Katz
Nathan Keller
Eike Kiltz
Jaeheon Kim
Matthias Kleinmann
Hugo Krawczyk
Konrad Kulikowski
Soonhak Kwon
Taekyoung Kwon
Tanja Lange
Cédric Lauradoux
Dong Hoon Lee
Hyang-Sook Lee
Anja Lehmann
Gaëtan Leurent
Pierre Loidreau
Mira Meyerovich
Marine Minier
Tal Moran
Sean Murphy
Christophe Nègre
Gregory Neven
Antonio Nicolosi
Kobbi Nissim
Shien Jin Ong
Yossi Oren
Raphael Overbeck

Adriana Palacio
Omkant Pandey
Alan Park
Rafael Pass
Michael Østergaard Pedersen
Chris Peikert
Ludovic Perret
Duong Hieu Phan
Krzysztof Pietrzak
Gilles Piret
David Pointcheval
Manoj Prabhakharan
Bartosz Przydatek
Charles W. Rackoff
Håvard Raddum
Mario Di Raimondo
Dominik Raub
Christian Rechberger
Omer Reingold
Leonid Reyzin
Thomas Ristenpart
Maike Ritzenhofen
Phil Rogaway
Amit Sahai
Louis Salvail
Berry Schoenmakers
Dominique Schröder
Gil Segev
Nicolas Sendrier
Emily Shen
Peter Shor
Alice Silverberg
William Speirs
François-Xavier Standaert
Damien Stehlé
Andreas Stein
Lakshminarayanan Subramanian
Madhu Sudan
Qiang Tang
Thomas Toft
Vinod Vaikuntanathan
Mayank Varia
Carmine Ventre
Ivan Visconti
David Wagner
Samuel Wagstaff
Shabsi Walfish
Bogdan Warinschi
Ralf-Philipp Weinmann
Enav Weinreb
Douglas Wikstrom
Jürg Wullschleger
Hyojin Yoon
Aram Yun

Lecture Notes in Computer Science

For information about Vols. 1–4381

please contact your bookseller or Springer

Vol. 4515: M. Naor (Ed.), Advances in Cryptology - EUROCRYPT 2007. XIII, 591 pages. 2007.

Vol. 4510: P. Van Hentenryck, L. Wolsey (Eds.), Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. X, 391 pages. 2007.

Vol. 4506: D. Zeng, I. Gotham, K. Komatsu, C. Lynch, M. Thurmond, D. Madigan, B. Lober, J. Kvach, H. Chen (Eds.), Intelligence and Security Informatics: Biosurveillance. XI, 234 pages. 2007.

Vol. 4504: J. Huang, R. Kowalczyk, Z. Maamar, D. Martin, I. Müller, S. Stoutenburg, K.P. Sycara (Eds.), Service-Oriented Computing: Agents, Semantics, and Engineering. X, 175 pages. 2007.

Vol. 4493: D. Liu, S. Fei, Z. Hou, H. Zhang, C. Sun (Eds.), Advances in Neural Networks – ISNN 2007, Part III. XXVI, 1215 pages. 2007.

Vol. 4492: D. Liu, S. Fei, Z. Hou, H. Zhang, C. Sun (Eds.), Advances in Neural Networks – ISNN 2007, Part II. XXVII, 1321 pages. 2007.

Vol. 4491: D. Liu, S. Fei, Z.-G. Hou, H. Zhang, C. Sun (Eds.), Advances in Neural Networks – ISNN 2007, Part I. LIV, 1365 pages. 2007.

Vol. 4486: M. Bernardo, J. Hillston (Eds.), Formal Methods for Performance Evaluation. VII, 469 pages. 2007.

Vol. 4484: J.-Y. Cai, S.B. Cooper, H. Zhu (Eds.), Theory and Applications of Models of Computation. XIII, 772 pages. 2007.

Vol. 4483: C. Baral, G. Brewka, J. Schlipf (Eds.), Logic Programming and Nonmonotonic Reasoning. IX, 327 pages. 2007. (Sublibrary LNAI).

Vol. 4482: A. An, J. Stefanowski, S. Ramanna, C.J. Butz, W. Pedrycz, G. Wang (Eds.), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. XIV, 585 pages. 2007. (Sublibrary LNAI).

Vol. 4481: J.T. Yao, P. Lingras, W.-Z. Wu, M. Szczuka, N.J. Cercone, D. Ślęzak (Eds.), Rough Sets and Knowledge Technology. XIV, 576 pages. 2007. (Sublibrary LNAI).

Vol. 4480: A. LaMarca, M. Langheinrich, K.N. Truong (Eds.), Pervasive Computing. XIII, 369 pages. 2007.

Vol. 4479: I.F. Akyildiz, R. Sivakumar, E. Ekici, J.C.d. Oliveira, J. McNair (Eds.), NETWORKING 2007. Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet. XXVII, 1252 pages. 2007.

Vol. 4472: M. Haindl, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems. XI, 524 pages. 2007.

Vol. 4471: P. Cesar, K. Chorianopoulos, J.F. Jensen (Eds.), Interactive TV: a Shared Experience. XIII, 236 pages. 2007.

Vol. 4470: Q. Wang, D. Pfahl, D.M. Raffo (Eds.), Software Process Change – Meeting the Challenge. XI, 346 pages. 2007.

Vol. 4464: E. Dawson, D.S. Wong (Eds.), Information Security Practice and Experience. XIII, 361 pages. 2007.

Vol. 4463: I. Măndoiu, A. Zelikovsky (Eds.), Bioinformatics Research and Applications. XV, 653 pages. 2007. (Sublibrary LNBI).

Vol. 4462: D. Sauveron, K. Markantonakis, A. Bilas, J.-J. Quisquater (Eds.), Information Security Theory and Practices. XII, 255 pages. 2007.

Vol. 4459: C. Cérin, K.-C. Li (Eds.), Advances in Grid and Pervasive Computing. XVI, 759 pages. 2007.

Vol. 4453: T. Speed, H. Huang (Eds.), Research in Computational Molecular Biology. XVI, 550 pages. 2007. (Sublibrary LNBI).

Vol. 4452: M. Fasli, O. Shehory (Eds.), Agent-Mediated Electronic Commerce. VIII, 249 pages. 2007. (Sublibrary LNAI).

Vol. 4450: T. Okamoto, X. Wang (Eds.), Public Key Cryptography – PKC 2007. XIII, 491 pages. 2007.

Vol. 4448: M. Giacobini et al. (Ed.), Applications of Evolutionary Computing. XXIII, 755 pages. 2007.

Vol. 4447: E. Marchiori, J.H. Moore, J.C. Rajapakse (Eds.), Evolutionary Computation,Machine Learning and Data Mining in Bioinformatics. XI, 302 pages. 2007.

Vol. 4446: C. Cotta, J. van Hemert (Eds.), Evolutionary Computation in Combinatorial Optimization. XII, 241 pages. 2007.

Vol. 4445: M. Ebner, M. O'Neill, A. Ekárt, L. Vanneschi, A.I. Esparcia-Alcázar (Eds.), Genetic Programming. XI, 382 pages. 2007.

Vol. 4444: T. Reps, M. Sagiv, J. Bauer (Eds.), Program Analysis and Compilation, Theory and Practice. X, 361 pages. 2007.

Vol. 4443: R. Kotagiri, P.R. Krishna, M. Mohania, E. Nantajeewarawat (Eds.), Advances in Databases: Concepts, Systems and Applications. XXI, 1126 pages. 2007.

Vol. 4440: B. Liblit, Cooperative Bug Isolation. XV, 101 pages. 2007.

Vol. 4439: W. Abramowicz (Ed.), Business Information Systems. XV, 654 pages. 2007.

Vol. 4438: L. Maicher, A. Sigel, L.M. Garshol (Eds.), Leveraging the Semantics of Topic Maps. X, 257 pages. 2007. (Sublibrary LNAI).

Vol. 4433: E. Şahin, W.M. Spears, A.F.T. Winfield (Eds.), Swarm Robotics. XII, 221 pages. 2007.

Vol. 4432: B. Beliczynski, A. Dzielinski, M. Iwanowski, B. Ribeiro (Eds.), Adaptive and Natural Computing Algorithms, Part II. XXVI, 761 pages. 2007.

Vol. 4431: B. Beliczynski, A. Dzielinski, M. Iwanowski, B. Ribeiro (Eds.), Adaptive and Natural Computing Algorithms, Part I. XXV, 851 pages. 2007.

Vol. 4430: C.C. Yang, D. Zeng, M. Chau, K. Chang, Q. Yang, X. Cheng, J. Wang, F.-Y. Wang, H. Chen (Eds.), Intelligence and Security Informatics. XII, 330 pages. 2007.

Vol. 4429: R. Lu, J.H. Siekmann, C. Ullrich (Eds.), Cognitive Systems. X, 161 pages. 2007. (Sublibrary LNAI).

Vol. 4427: S. Uhlig, K. Papagiannaki, O. Bonaventure (Eds.), Passive and Active Network Measurement. XI, 274 pages. 2007.

Vol. 4426: Z.-H. Zhou, H. Li, Q. Yang (Eds.), Advances in Knowledge Discovery and Data Mining. XXV, 1161 pages. 2007. (Sublibrary LNAI).

Vol. 4425: G. Amati, C. Carpineto, G. Romano (Eds.), Advances in Information Retrieval. XIX, 759 pages. 2007.

Vol. 4424: O. Grumberg, M. Huth (Eds.), Tools and Algorithms for the Construction and Analysis of Systems. XX, 738 pages. 2007.

Vol. 4423: H. Seidl (Ed.), Foundations of Software Science and Computational Structures. XVI, 379 pages. 2007.

Vol. 4422: M.B. Dwyer, A. Lopes (Eds.), Fundamental Approaches to Software Engineering. XV, 440 pages. 2007.

Vol. 4421: R. De Nicola (Ed.), Programming Languages and Systems. XVII, 538 pages. 2007.

Vol. 4420: S. Krishnamurthi, M. Odersky (Eds.), Compiler Construction. XIV, 233 pages. 2007.

Vol. 4419: P.C. Diniz, E. Marques, K. Bertels, M.M. Fernandes, J.M.P. Cardoso (Eds.), Reconfigurable Computing: Architectures, Tools and Applications. XIV, 391 pages. 2007.

Vol. 4418: A. Gagalowicz, W. Philips (Eds.), Computer Vision/Computer Graphics Collaboration Techniques. XV, 620 pages. 2007.

Vol. 4416: A. Bemporad, A. Bicchi, G. Buttazzo (Eds.), Hybrid Systems: Computation and Control. XVII, 797 pages. 2007.

Vol. 4415: P. Lukowicz, L. Thiele, G. Tröster (Eds.), Architecture of Computing Systems - ARCS 2007. X, 297 pages. 2007.

Vol. 4414: S. Hochreiter, R. Wagner (Eds.), Bioinformatics Research and Development. XVI, 482 pages. 2007. (Sublibrary LNBI).

Vol. 4412: F. Stajano, H.J. Kim, J.-S. Chae, S.-D. Kim (Eds.), Ubiquitous Convergence Technology. XI, 302 pages. 2007.

Vol. 4411: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 249 pages. 2007. (Sublibrary LNAI).

Vol. 4410: A. Branco (Ed.), Anaphora: Analysis, Algorithms and Applications. X, 191 pages. 2007. (Sublibrary LNAI).

Vol. 4409: J.L. Fiadeiro, P.-Y. Schobbens (Eds.), Recent Trends in Algebraic Development Techniques. VII, 171 pages. 2007.

Vol. 4407: G. Puebla (Ed.), Logic-Based Program Synthesis and Transformation. VIII, 237 pages. 2007.

Vol. 4406: W. De Meuter (Ed.), Advances in Smalltalk. VII, 157 pages. 2007.

Vol. 4405: L. Padgham, F. Zambonelli (Eds.), Agent-Oriented Software Engineering VII. XII, 225 pages. 2007.

Vol. 4403: S. Obayashi, K. Deb, C. Poloni, T. Hiroyasu, T. Murata (Eds.), Evolutionary Multi-Criterion Optimization. XIX, 954 pages. 2007.

Vol. 4401: N. Guelfi, D. Buchs (Eds.), Rapid Integration of Software Engineering Techniques. IX, 177 pages. 2007.

Vol. 4400: J.F. Peters, A. Skowron, V.W. Marek, E. Orłowska, R. Słowiński, W. Ziarko (Eds.), Transactions on Rough Sets VII, Part II. X, 381 pages. 2007.

Vol. 4399: T. Kovacs, X. Llorà, K. Takadama, P.L. Lanzi, W. Stolzmann, S.W. Wilson (Eds.), Learning Classifier Systems. XII, 345 pages. 2007. (Sublibrary LNAI).

Vol. 4398: S. Marchand-Maillet, E. Bruno, A. Nürnberger, M. Detyniecki (Eds.), Adaptive Multimedia Retrieval: User, Context, and Feedback. XI, 269 pages. 2007.

Vol. 4397: C. Stephanidis, M. Pieper (Eds.), Universal Access in Ambient Intelligence Environments. XV, 467 pages. 2007.

Vol. 4396: J. García-Vidal, L. Cerdà-Alabern (Eds.), Wireless Systems and Mobility in Next Generation Internet. IX, 271 pages. 2007.

Vol. 4395: M. Daydé, J.M.L.M. Palma, Á.L.G.A. Coutinho, E. Pacitti, J.C. Lopes (Eds.), High Performance Computing for Computational Science - VECPAR 2006. XXIV, 721 pages. 2007.

Vol. 4394: A. Gelbukh (Ed.), Computational Linguistics and Intelligent Text Processing. XVI, 648 pages. 2007.

Vol. 4393: W. Thomas, P. Weil (Eds.), STACS 2007. XVIII, 708 pages. 2007.

Vol. 4392: S.P. Vadhan (Ed.), Theory of Cryptography. XI, 595 pages. 2007.

Vol. 4391: Y. Stylianou, M. Faundez-Zanuy, A. Esposito (Eds.), Progress in Nonlinear Speech Processing. XII, 269 pages. 2007.

Vol. 4390: S.O. Kuznetsov, S. Schmidt (Eds.), Formal Concept Analysis. X, 329 pages. 2007. (Sublibrary LNAI).

Vol. 4389: D. Weyns, H.V.D. Parunak, F. Michel (Eds.), Environments for Multi-Agent Systems III. X, 273 pages. 2007. (Sublibrary LNAI).

Vol. 4385: K. Coninx, K. Luyten, K.A. Schneider (Eds.), Task Models and Diagrams for Users Interface Design. XI, 355 pages. 2007.

Vol. 4384: T. Washio, K. Satoh, H. Takeda, A. Inokuchi (Eds.), New Frontiers in Artificial Intelligence. IX, 401 pages. 2007. (Sublibrary LNAI).

Vol. 4383: E. Bin, A. Ziv, S. Ur (Eds.), Hardware and Software, Verification and Testing. XII, 235 pages. 2007.

¥775.00元

Table of Contents

Chosen-Prefix Collisions for MD5 and Colliding X.509 Certificates for Different Identities

Marc Stevens[1], Arjen Lenstra[2], and Benne de Weger[1]

[1] TU Eindhoven, Faculty of Mathematics and Computer Science
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[2] EPFL IC LACAL, Station 14, and Bell Laboratories
CH-1015 Lausanne, Switzerland

Abstract. We present a novel, automated way to find differential paths for MD5. As an application we have shown how, at an approximate expected cost of 2^{50} calls to the MD5 compression function, for any two chosen message prefixes P and P', suffixes S and S' can be constructed such that the concatenated values $P\|S$ and $P'\|S'$ collide under MD5. Although the practical attack potential of this construction of *chosen-prefix collisions* is limited, it is of greater concern than random collisions for MD5. To illustrate the practicality of our method, we constructed two MD5 based X.509 certificates with identical signatures but different public keys *and* different Distinguished Name fields, whereas our previous construction of colliding X.509 certificates required identical name fields. We speculate on other possibilities for abusing chosen-prefix collisions. More details than can be included here can be found on `www.win.tue.nl/hashclash/ChosenPrefixCollisions/`.

1 Introduction

In March 2005 we showed how Xiaoyun Wang's ability [17] to quickly construct random collisions for the MD5 hash function could be used to construct two different valid and unsuspicious X.509 certificates with identical digital signatures (see [10] and [11]). These two *colliding certificates* differed in their public key values only. In particular, their Distinguished Name fields containing the identities of the certificate owners were equal. This was the best we could achieve because

- Wang's hash collision construction requires identical Intermediate Hash Values (IHVs);
- the resulting colliding values look like random strings: in an X.509 certificate the public key field is the only suitable place where such a value can unsuspiciously be hidden.

A natural and often posed question (cf. [7], [3], [1]) is if it would be possible to allow more freedom in the other fields of the certificates, at a cost lower than 2^{64}

M. Naor (Ed.): EUROCRYPT 2007, LNCS 4515, pp. 1–22, 2007.

calls to the MD5 compression function. Specifically, it has often been suggested that it would be interesting to be able to select Distinguished Name fields that are different and, preferably, chosen at will, non-random and human readable as one would expect from these fields. This can be realized if two arbitrarily chosen messages, resulting in two different IHVs, can be extended in such a way that the extended messages collide. Such collisions will be called *chosen-prefix collisions.*

We describe how chosen-prefix collisions for MD5 can be constructed, and show that our method is practical by constructing two MD5 based X.509 certificates with different Distinguished Name fields and identical digital signatures. The full details of the chosen-prefix collision construction and the certificates can be found in [16] and [14], respectively.

Section 2 contains a bird's eye view of the chosen-prefix collision construction method and its complexity. Its potential applications are discussed in Section 3 with Section 4 containing implications and details of the application to X.509 certificates. Details of the automated differential path construction for MD5 are provided in Section 5.

2 Chosen-Prefix Collisions for MD5

The main contribution of this paper is a method to construct MD5 collisions starting from two arbitrary IHVs. Given this method one can take any two chosen message prefixes and construct bitstrings that, when appended to the prefixes, turn them into two messages that collide under MD5. We refer to such a collision as a *chosen-prefix collision.* Their possibility was mentioned already in [3, Section 4.2 case 1] and, in the context of SHA-1, in [1] and on `www.iaik.tugraz.at/research/krypto/collision/`.

We start with a pair of arbitrarily chosen messages, not necessarily of the same length. Padding with random bits may be applied so that the padded messages have the same bitlength which equals 416 modulo 512 (incomplete last block). Equal length is unavoidable, because Merkle-Damgård strengthening, involving the message length, is applied after the last message block has been compressed by MD5. The incomplete last block condition is a technical requirement. In our example of colliding certificates the certificate contents were constructed in such a way that padding was not necessary, to allow for shorter RSA moduli.

Given the padded message pair, we followed a suggestion by Xiaoyun Wang[1] to find a pair of 96-bit values that, when used to complete the last blocks by appending them to the messages and applying the MD5 compression function, resulted in a specific form of difference vector between the IHVs. Finding these 96-bit values was done using a birthdaying procedure.

The remaining differences between the IHVs were then removed by appending *near-collision blocks.* Per pair of blocks this was done by constructing new differential paths using an automated, improved version of Wang's original approach. This innovative differential path construction is described in detail in Section 5

[1] Private communication.

below. Due to the specific form of the near-collisions and the first difference vector, essentially one triple of bit differences could be removed per near-collision block, thus shortening the overall length of the colliding values. For our example 8 near-collision blocks were needed to remove all differences. Thus, a total of $96 + 8 \times 512 = 4192$ bits were appended to each of the chosen message prefixes to let them collide.

The birthdaying step can be entirely avoided, thereby making it harder to find the proper differential paths and considerably increasing the number of near-collision blocks. Or the birthdaying step could be simplified, increasing the number of near-collision blocks from 8 to about 14. Our approach was inspired by our desire to minimize the number of near-collision blocks. Using a more intricate differential path construction it should be possible to remove more than a single triple of bit differences per block, which would reduce the number of near-collision blocks. Potential enhancements and variations, and the full details of the construction as used, will be discussed in [16].

The expected complexity of the birthdaying step is estimated at 2^{49} MD5 compression function calls. Estimating the complexity of the near-collision block construction is hard, but it turned out to be a small fraction of the birthdaying complexity. Based on our observations we find it reasonable to estimate the overall expected complexity of finding a chosen-prefix collision for MD5 at about 2^{50} MD5 compression function calls. For the example we constructed, however, we had some additional requirements and also were rather unlucky in the birthdaying step, leading to about 2^{52} MD5 compression function calls. Note that, either way, this is substantially faster than the trivial birthday attack which has complexity 2^{64}.

The construction of just a single example required, apart from the development of the automated differential path construction method, substantial computational efforts. Fortunately, the work is almost fully parallelizable and suitable for grid computing. It was done in the "HashClash" project (see `www.win.tue.nl/hashclash/`) and lasted about 6 months: using BOINC software (see `boinc.berkeley.edu/`) up to 1200 machines contributed, involving a cluster of computers at TU/e and a grid of home PCs. We expect that another chosen-prefix collision can be found much faster, but that it would again require substantial effort, both human and computationally: say 2 months real time assuming comparable computational resources.

3 Applications of Chosen-Prefix Collisions

We mention some potential applications of chosen-prefix collisions.

- The example presented in the next section, namely colliding X.509 certificates with different fields before the appended bitstrings that cause the collision. Those bitstrings are 'perfectly' hidden inside the RSA moduli, where 'perfect' means that inspection of either one of the RSA moduli does not give away anything about the way it is constructed (namely, crafted such