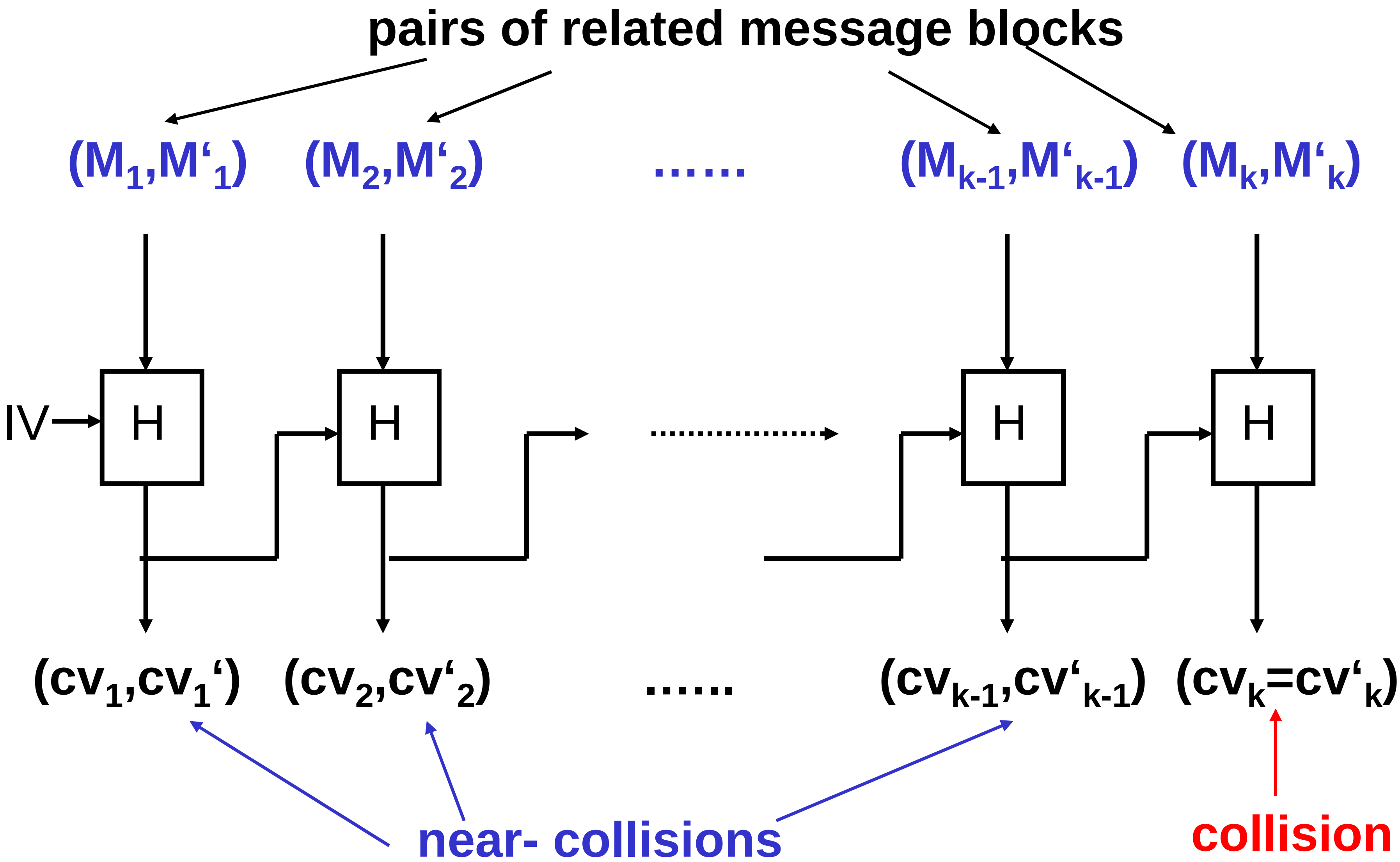
Precise Probabilities for Hash Collision Paths

M. Gebhardt, G. Illies, W. Schindler

Bundesamt für Sicherheit in der Informationstechnik (BSI), Bonn

Santa Barbara, August 25, 2006

(Multiblock) Hash Collision Attack



Workload for a collision

- The pair (cv_j,cv'_j) is called a near-collision if both components are „almost“ equal, fulfilling a set of specified conditions.

- Workload = Workload (Block1) + ... + Workload (Block k)
  Consequence: The blocks may be analysed independently.

- Set of sufficient (bit) conditions *SC*
  - characterizes a (near-)collision path
  - $\rightarrow$ (near-)collision

Workload and Success Probability

- *SC = SC1* $\cup$ *SC2*
  - *SC1*: conditions can be guaranteed by message modification
  - *SC2* (conditions after message modification): fulfilled with a particular probability

- Prob(near-collision path) = Prob(all SC2-conditions are fulfilled)
  - Prob((near-)collision) $\geq$ Prob((near-) collision path)
  - $\rightarrow$ workload

The set *SC2*

- Example

  $SC2 := \{(r_{27,5}, r'_{27,5}) = (0,1), r_{34,5} = r_{33,5}, (r_{45,25}, r'_{45,25}) = (0,0), \ldots\}$

  where $r_{i,j}$ = register bit j in Step i

- „Rule of thumb“ (usually applied):

  Prob(near-collision path) =

  Prob(all cond's. of *SC2* are fulfilled) $\approx 2^{-|SC2|}$

  number of bit conditions

Goal of this contribution

- This rule of thumb provides only a rough estimate of the true probabilities.
- Deviations may be caused by various interfering effects:
  - cyclical shifts
  - addition of 32-bit words ($\rightarrow$ carry bits)
  - bit conditions on the chaining values (post addition with fixed values; bit counting is very inaccurate)

  NOTE: Specific effects have been addressed in literature (qualitatively and / or quantitatively)
- Our contribution supplies universal tools that support the systematic calculation of probabilities of (near-)collision paths.

Stochastic Model

- Step functions (examples)
  - (MD5) $r_i = r_{i-1} + (\ \Phi_i(r_{i-1},r_{i-2},r_{i-3}) + r_{i-4} + m_i + const_i\ )^{<<<s} \pmod{2^{32}}$
  - (SHA-1) $r_i = r_{i-1}^{<<<5} + \Phi_i(r_{i-2},r_{i-3},r_{i-4}) + r_{i-5} + m_i + const_i \pmod{2^{32}}$

    $r_{i-2} = r_{i-2}^{<<<30}$

- Stochastic model

  We interpret the intermediate register values $(r_1,r'_1), (r_2,r'_2),\ldots$ and the message blocks $(m_1,m'_1), (m_2,m'_2),\ldots$ as values assumed by random variables $(R_1,R'_1), (R_2,R'_2),\ldots$ and $(M_1,M'_1), (M_2,M'_2), \ldots$, respectively.

  These random variables have specific properties which depend on the hash function and the near-collision path.

Relevant Types of Probabilities

- Notation:
  - The random variables X, X', Y, Y' assume values in $Z_{2^{32}}$
  - $S_1, S_2, S_3 \subseteq Z_{2^{32}} \times Z_{2^{32}}$ denote specific subsets ($\rightarrow$ bit conditions)
  - $T_i := pr_1(S_i) \subseteq Z_{2^{32}}$ (projection onto the 1st component)

- Relevant types of conditional probabilities:
  - $\mathrm{Prob}((X,X') + (Y,Y') \pmod{2^{32}} \in S_3 \mid (X,X') \in S_1, (Y,Y') \in S_2)$
  - $\mathrm{Prob}((X,X')^{<<<s} + (Y,Y') \pmod{2^{32}} \in S_3 \mid (X,X') \in S_1, (Y,Y') \in S_2)$
  - $\mathrm{Prob}((X,X')^{<<<s} + (Y,Y') \pmod{2^{32}} \in S_3 \mid (X-X') \pmod{2^{32}} = \Delta, (Y,Y') \in S_2)$

Main results

- Under suitable assumptions the conditional probabilities from the last slide can be simplified to
  - $\mathrm{Prob}(X+Y \pmod{2^{32}} \in T_3 \mid X \in T_1, Y \in T_2) * 1_{\{0\}}(A[S_1,S_2,S_3])$
  - $\mathrm{Prob}(X^{<<<s} + Y \pmod{2^{32}} \in T_3 \mid X \in T_1, Y \in T_2) * 1_{\{0\}}(B[s,S_1,S_2,S_3])$
  - $\mathrm{Prob}(X^{<<<s} + Y \pmod{2^{32}} \in T_3 \mid X \in V[s,S_1,S_2,S_3], Y \in T_2) * \mathrm{Prob}(X \in V[s,S_1,S_2,S_3])$

The paper provides characterisations for the conditions $A[S_1,S_2,S_3]$, $B[s,S_1,S_2,S_3]$ and for the set $V[s,S_1,S_2,S_3]$ that are appropriate for concrete calculations.

Example: MD5, Block 1 (1)

Stochastic model: $\rightarrow$ paper

Impact of bit conditions on the chaining values:

Post additions in Steps 61- 63: 6 bit conditions

- Wang Conditions (Eurocrypt 2005, PAPER):
  - Transition probability for standard IV $\approx$ 0.005

- Wang Conditions (Eurocrypt 2005, PUBLISHED EXAMPLE):
  - Transition probability for standard IV $\approx$ 0.095
  - Transition probability for IV = (0x 80000000, 0x 00000000, 0x 82000000, 0x 10325476) = 0.5
  - Transition probability for IV=(0x 00000000, 0x 82000000, 0x 80000000, 0x 10325476) = 0

Example: MD5, Block 1 (2)

- We analysed three different near-collision paths after message modification:
  - Path 1: Wang Conditions (PAPER, Eurocrypt 2005)
  - Path 2: Wang Conditions (PUBLISHED EXAMPLE)
  - Path 3: “Almost”-Wang conditions

	Path1	Path 2	Path3
# bit conditions	38	38	39
calculated probability	$2^{-41.64}$	$2^{-37.41}$	$2^{-36.61}$
empirical ($2^{41.87}$ samples)		$2^{-37.11}$	$2^{-36.25}$

Conclusion

- “Bit condition counting” yields only rough estimators for the probabilities of (near-)collision paths.
- Our contribution provides universally applicable theorems that support the precise computation of collision path probabilities.
- These theorems do not support the search for new (near-) collision paths.
- Our formulae were empirically confirmed by concrete MD5 near-collision paths.

Contact

Bundesamt für Sicherheit in der Informationstechnik (BSI)

Werner Schindler
Godesberger Allee 185-189
53175 Bonn
Germany

Tel: +49 (0)1888-9582-5652
Fax: +49 (0)1888-10-9582-5652

Werner.Schindler@bsi.bund.de
www.bsi.bund.de
www.bsi-fuer-buerger.de