Linear Analysis of Reduced-Round Skein

Tomer Ashur

Faculty of Mathematics and Computer Science
Weizmann Institute of Science
tomerashur@gmail.com

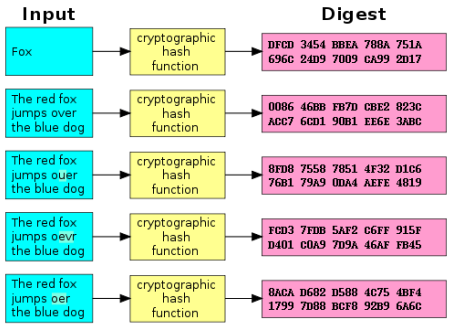
19/6/11

NIST SHA-3 Competition

- In 2007 NIST has announced a public competition for adding a new hash-function to the SHA family.
- 64 submissions, 51 candidates, 14 in the second round, 5 in the last round.
- Amongst which was Skein by Niels Ferguson, Stefan Lucks, Bruce Schneier, Doug Whiting, Mihir Bellare, Tadayoshi Kohno, Jon Callas and Jesse Walker.

Brief Intorduction to Hash Functions

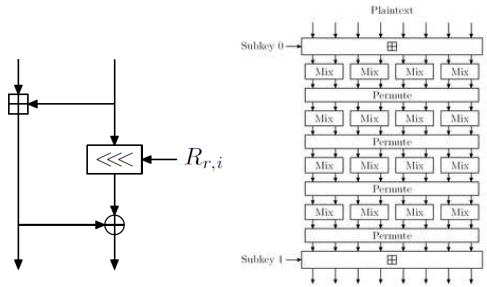
- A function that maps inputs of arbitrary length into an output of fixed size.
- Widely used hash functions include: MD5, SHA1, SHA-256, WHIRLPOOL and others.
- Used as a primitive in many cryptographic protocols (signature schemes, authentication schemes, PRNGs etc.).



By User:Jorge Stolfi based on Image:Hash function.svg by Helix84 (Original work for Wikipedia) [Public domain], via Wikimedia Commons

Structure of Skein

- Skein has an internal state of 256, 512 or 1024 bits.
- The compression function is based on the Threefish block cipher
- Threefish is applied on the state iteratively.



Both images are by bruce schneier et al. (Skein paper v1.3) [GPL (www.gnu.org/licenses/gpl.html) or CC0 (creativecommons.org/publicdomain/zero/1.0/deed.en)], via Wikimedia Commons

Linear Cryptanalysis

- Linear cryptanalysis is a useful cryptanalytic tool to attack block ciphers.
- A linear cryptanalysis attack has two parts: finding a linear approximation and using a linear approximation to attack the cryptosystem.

Finding a Good Linear Approximation

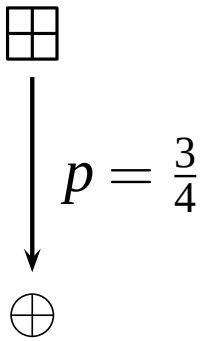
- The adversary tries to approximate the nonlinear operations with other (linear) operations.
- The result is an expression of the form $P_i \oplus \ldots \oplus P_j \oplus C_k \oplus \ldots \oplus C_l = K_m \oplus \ldots \oplus K_n$.
  - The P's are bits from the plaintext, the C's are bits from the ciphertext and the K's are bits from the key.
- Each approximation has a probability p assosicated with it.

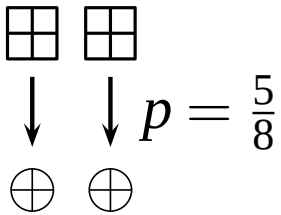
Linear Attack

- Once an approximation is found, the adversary can use it to recover bits of the key or to distinguish the function from a random one.
- The adversary asks for pairs of input and output and evaluates the expression using them.
- Each such successful evaluation gives 1-bit worth of information about the key.

Linear Approximation of Modular Addition

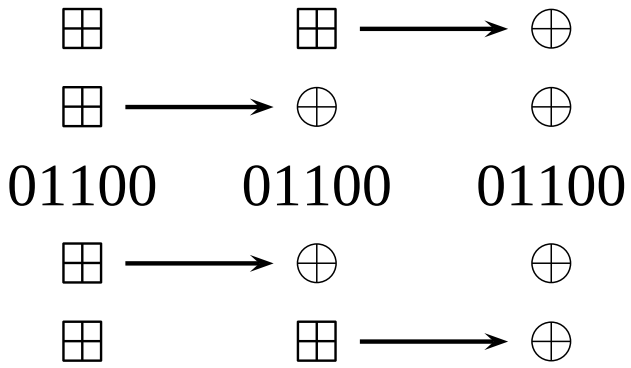
- The only nonlinear operation in CubeHash is the addition modulo 2^{64}.
- However, two consecutive bits entering an addition give rise to a bias of $\frac{1}{4}$ in the output. [Cho and Pieperzyk]
- Moreover, any even number of consecutive bits can be handled as an independant pair (i.e., 4 consecutive bits can be treated as 2 pairs of consecutive bits).





Our Work

- Using a C program we iterated all pairs of consecutive bits when running the round-function both forward and backward.
- The iteration for a certain pair stops when one of these events occured:
  - The rotation operation sent a pair of approximated bits to the MSB and LSB hence not adhering to the Cho and Pieperzyk framework.
  - A XOR operation create a single bit (i.e., $11, 12 \oplus 12, 13 = 11, 13$) hence not adhering to the Cho and Pieperzyk framework.
  - The total bias has become smaller than 2^{-256}.



$$0110 \lll 2 = 1001$$

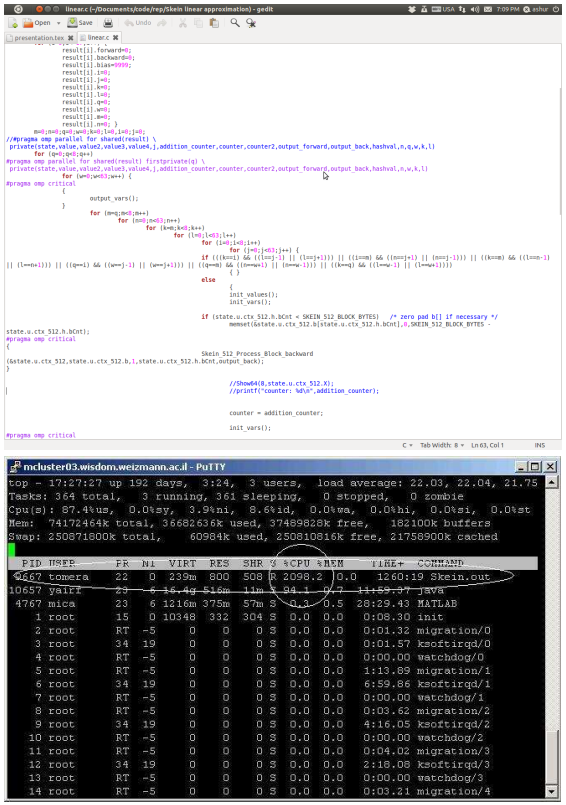
$$0110 \oplus 0011 = 0101$$

Computational Challenge

- Our goal is to search a space which is as large as possible.
- The size of the search space is increased as we search for more pairs.
- The complexity of the search:
  - Single pair: $\binom{1}{510} \approx 2^{9}$
  - Two pairs: $\binom{2}{510} \approx 2^{17}$
  - Three pairs: $\binom{3}{510} \approx 2^{24}$
  - Four pairs: $\binom{4}{510} \approx 2^{31}$
  - Five pairs: $\binom{4}{510} \approx 2^{38}$
- The code is in the form of nested ”for” loops, each additional pair adds another layer of loops.

Computational Solution

- Using OpenMPI I've parallelized the for loops.
- Parallelization is done to one of the inner loop in order to increase the use of the computer number of processors.
- This allows utilizing the full power of the WIS cluster server.



That's All Folks

Questions?