KU LEUVEN

Quasidifferential cryptanalysis of the MD5 hash function

Honours programme

Petar Vitorac

Faculty of Engineering Science

15 May 2024

Outline

1. Analysis of an example cipher
2. Description of MD5
3. Analysis of MD5
4. Future work
5. Conclusion

Outline

1. Analysis of an example cipher
   - Description
   - Probability of a characteristic
2. Description of MD5
3. Analysis of MD5
4. Future work
5. Conclusion

Example cipher

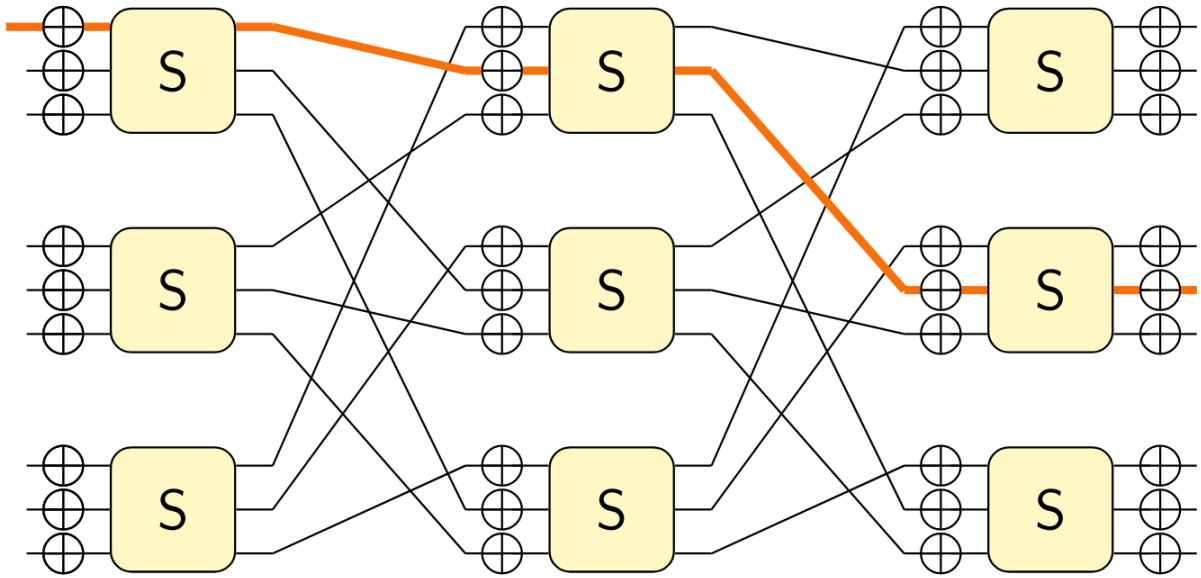


Figure: Example cipher

Probability of a characteristic

- Average probability: $(2^{-2})^3 = 2^{-6} \rightarrow$ no consideration for keys
- Quasidifferential transition matrix for the S-Box
- Find quasidifferential trails
- $2^{-6} + (-1)^{k_{2,8}+k_{3,4}} 2^{-7} + (-1)^{k_{2,5}+k_{3,6}} 2^{-7} + (-1)^{k_{2,8}+k_{3,4}+k_{2,5}+k_{3,6}} 2^{-8}$
- Example: when keys are zero: $9 \cdot 2^{-8}$

Outline

1. Analysis of an example cipher
2. Description of MD5
3. Analysis of MD5
4. Future work
5. Conclusion

Description of MD5

- Input: arbitrary length $\rightarrow$ 512-bit blocks
- Output: 128-bit hash
- Four 32-bit word states
- 64 steps split in 4 rounds

$$A \leftarrow B$$
$$B \leftarrow C$$
$$C \leftarrow D$$
$$D \leftarrow ((A + F_j(B, C, D) + M[g_j(i)] + K[i]) \lll S[i]) + B$$

- Different F-function for each round (if-else, xor, ...)

Outline

1. Analysis of an example cipher
2. Description of MD5
3. Analysis of MD5
   - Characteristics
   - Quasidifferential trails
   - Conditions on the message
4. Future work
5. Conclusion

Characteristics

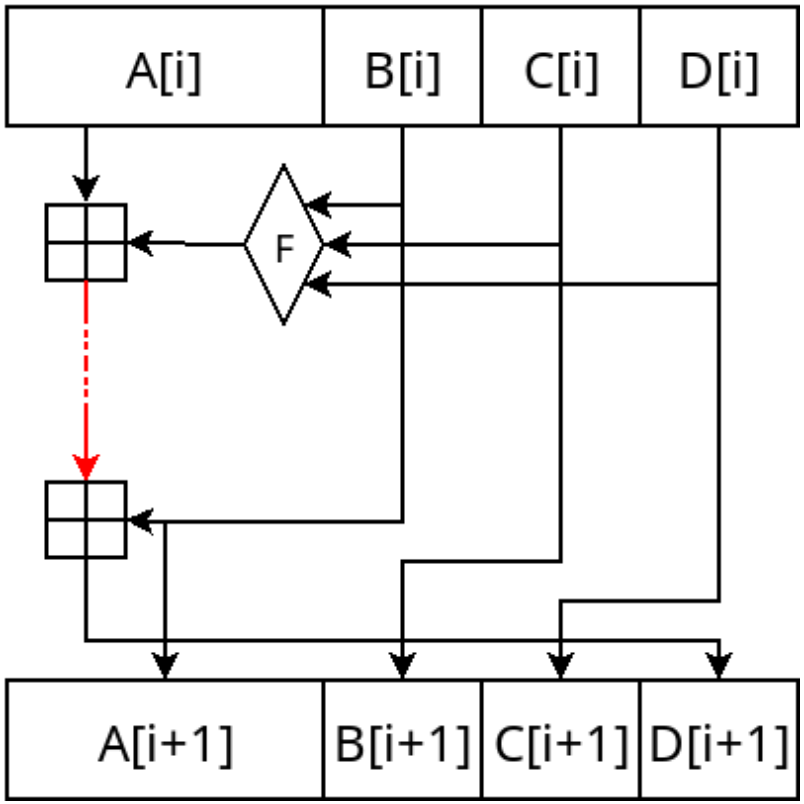


Figure: A step of MD5. Red represents undetermined differences.

Modelling

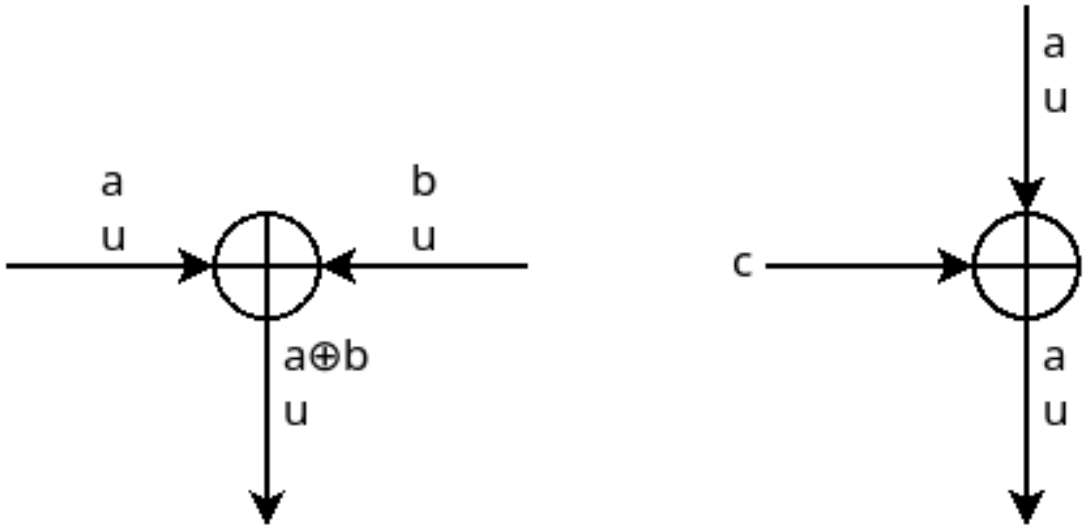


Figure: XOR

Modelling

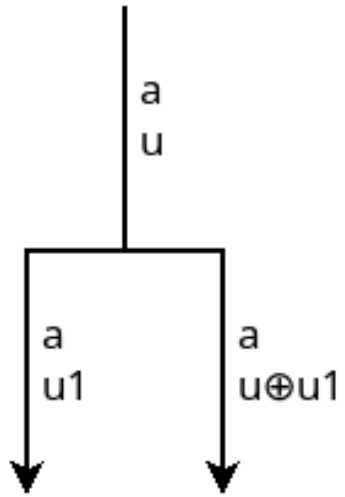
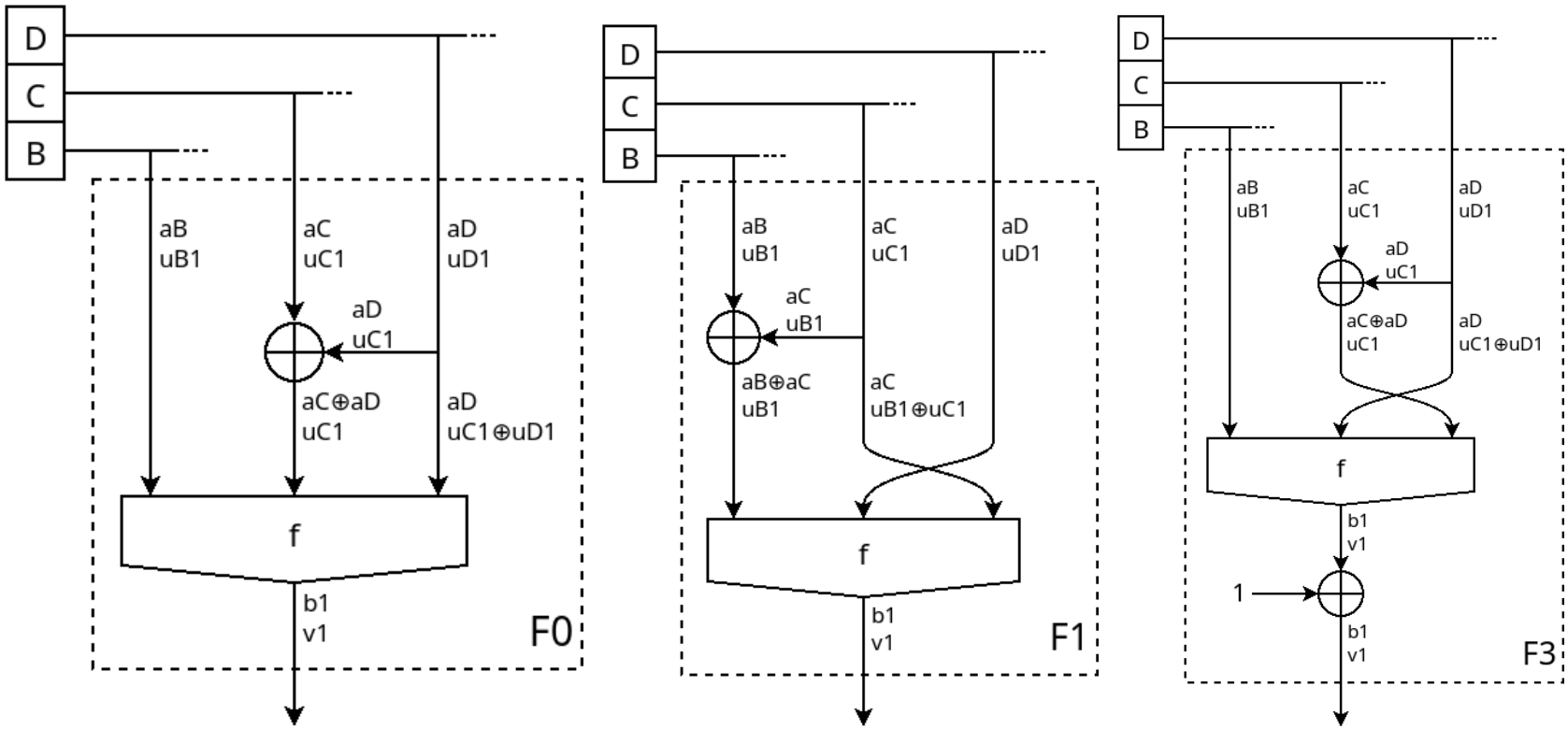


Figure: Branching

Modelling



$$f(x, y, z) = (x \wedge y) \oplus z$$

$$F_0(x, y, z) = (x \wedge y) \vee (x' \wedge z) \quad F_1(x, y, z) = (x \wedge z) \vee (y \wedge z')$$

$$F_2(x, y, z) = x \oplus y \oplus z \quad F_3(x, y, z) = y \oplus (x \vee z')$$

Modelling

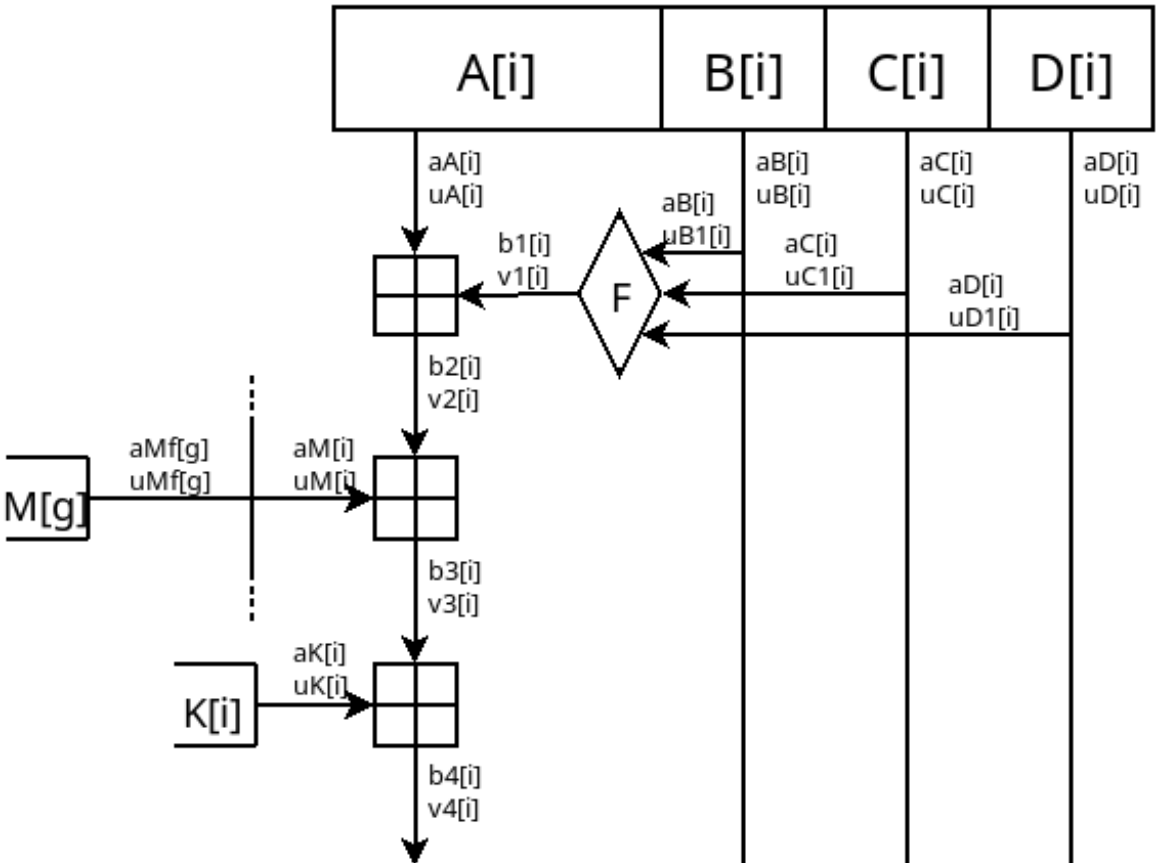


Figure: MD5 step (1)

Modelling

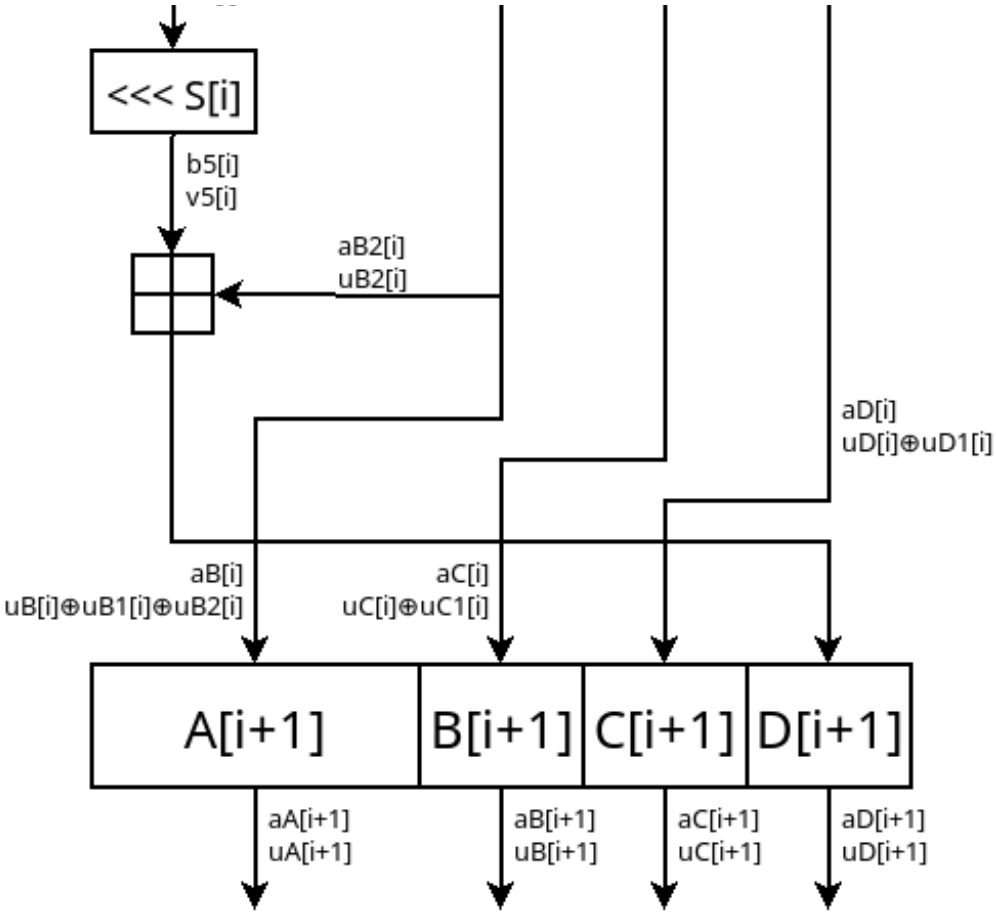


Figure: MD5 step (2)

Computing trails

- Satisfiability Modulo Theories
- *Boolector*
- Masks $\rightarrow$ unknown 32-bit vectors
- Assertions from the model
- Assertions for the weight

Filtering out solutions

- Unique solutions
- Linearly independent masks

Linearly independent masks

Example: $u_1 = [1\ 0\ 0\ 1], u_2 = [0\ 0\ 1\ 1]$
New mask: $u = [a\ b\ c\ d]$

$$\begin{bmatrix} 1 & 0 & a \\ 0 & 0 & b \\ 0 & 1 & c \\ 1 & 1 & d \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & c \\ 0 & 0 & b \\ 0 & 0 & a \oplus c \oplus d \end{bmatrix}$$

$$\rightarrow b = 1 \text{ or } a \oplus c \oplus d = 1$$

In general:
$[u_1^\top | u_2^\top | \ldots | u_n^\top] \rightarrow$ row echelon form via T
Last column is Tu

Conditions on the message

- Implicitly determined by trails
- $u^\top x = s$ with
  - u the input and output masks
  - x the input/output
  - $(-1)^s$ the sign of the trail
- $\rightarrow u_m^\top m = s^*$ with
  - u_m the mask on the message
  - m the message
- **Probabilistic**

Deterministic conditions

- Q.d. correlation = probability of differential $\times$ linear correlation
- Observation: many trails with weight same as that for zero mask
- "**Deterministic**" $\rightarrow$ linear relation always holds
- Multiple trails with the same message mask irrelevant

Results

36 (resp. 30) linearly independent conditions on two characteristics

Outline

1. Analysis of an example cipher
2. Description of MD5
3. Analysis of MD5
4. Future work
5. Conclusion

Future work

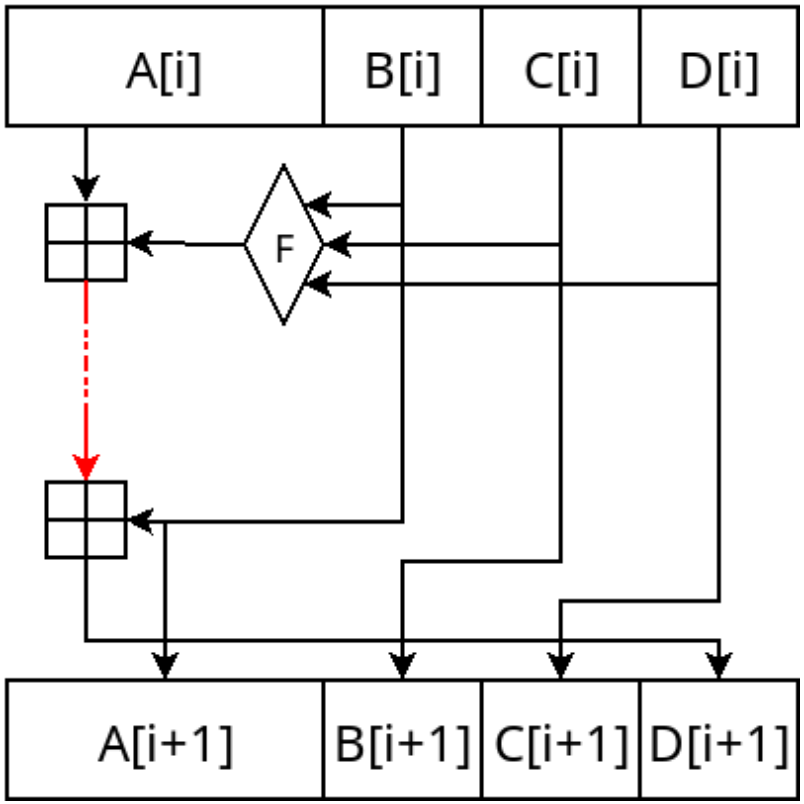


Figure: A step of MD5. Red represents undetermined differences.

Outline

1. Analysis of an example cipher
2. Description of MD5
3. Analysis of MD5
4. Future work
5. Conclusion

Conclusion

- Positive results
  - Applied quasidifferential methods to MD5
  - Reducing search space for collisions
- Possible improvements
  - Reconcile different characteristics
- Positive personal experience

Personal experience

- Learned a lot in a new domain
- Research skills
- Engineering competences
- Soft skills
- FSE conference
- Social relevance
- Learning from mistakes

https://github.com/petar-vitorac/md5-quasidifferential