

Хеширование строк

1 Хеш-функции

Хеш-функция предназначена для свертки входного массива любого размера в битовую строку, для MD5 длина выходной строки равна 128 битам. Для чего это нужно? К примеру у вас есть два массива, а вам необходимо быстро сравнить их на равенство, то хеш-функция может сделать это за вас, если у двух массивов хеши разные, то массивы гарантировано разные, а в случае равенства хешей — массивы скорее всего равны.

Хеш-функции используются в криптографических алгоритмах, электронных подписях, кодах аутентификации сообщений, обнаружении манипуляций, сканировании отпечатков пальцев, контрольных суммах (проверка целостности сообщений), хеш-таблицах, хранении паролей и многом другом. К примеру, скачивая файл из интернета, вы часто видите рядом с ним строку вида b10a8db164e0754105b7a99be72e3fe5 — это и есть хеш, прогнав этот файл через алгоритм MD5 вы получите такую строку, и, если хэши равны, можно с большой вероятностью утверждать что этот файл действительно подлинный (конечно с некоторыми оговорками, о которых расскажу далее). [1]

2 MD5

MD5 (англ. Message Digest 5) — 128-битный алгоритм хеширования, разработанный профессором Рональдом Л. Ривестом из Массачусетского технологического института (Massachusetts Institute of Technology, MIT) в 1991 году. Предназначен для создания «отпечатков» или дайджестов сообщения произвольной длины и последующей проверки их подлинности. Широко применялся для проверки целостности информации и хранения хешей паролей.

Алгоритм производит хеш со значением в 128 битов. Широко используется для проверки целостности данных. Не подходит для использования в иных областях по причине уязвимости в безопасности MD5. (рис.1)

Алгоритм состоит из пяти шагов:

1) Append Padding Bits

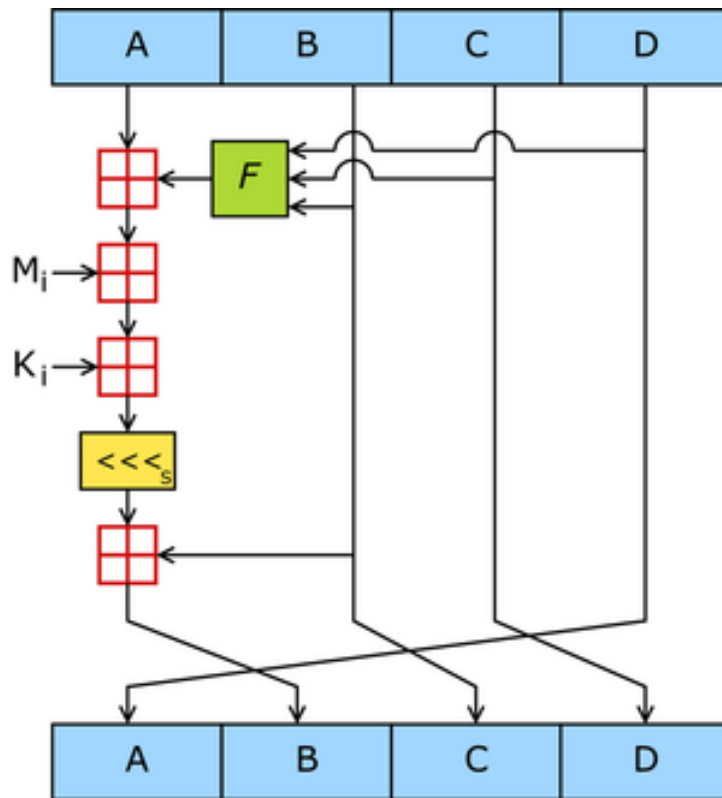


Рис. 1: Схема работы алгоритма MD5

В исходную строку дописывают единичный байт 0x80, а затем дописывают нулевые биты, до тех пор, пока длина сообщения не будет сравнима с 448 по модулю 512. То есть дописываем нули до тех пор, пока длина нового сообщения не будет равна $[длина] = (512 \cdot N + 448)$, где N — любое натуральное число, такое, что это выражение будет наиболее близко к длине блока.

2) Append Length

Далее в сообщение дописывается 64-битное представление длины исходного сообщения.

3) Initialize MD Buffer

На этом шаге инициализируется буффер

word A: 01 23 45 67

word B: 89 ab cd ef

word C: fe dc ba 98

word D: 76 54 32 10

Как можно заметить буффер состоит из четырех констант, предназначенный для сбора хэша.

4) Process Message in 16-Word Blocks

На четвертом шаге в первую очередь определяется 4 вспомогательные логические функции, которые преобразуют входные 32-битные слова, в, как ни странно, в 32-битные выходные.

5) Output

Выводя побайтово буффер ABCD начиная с A и заканчивая D получим наш хеш. В листинге 1 представлен пример MD5-хеширования.

Листинг 1: MD5

```
import hashlib
hash_object = hashlib.md5(b'Hello World')
print(hash_object.hexdigest())
```

3 SHA-1

Default3.

Список литературы

- [1] Марков А. С. and Цирлов В. Л. *Спецификация системы управления информационной безопасностью*. APS, 2021.