академия больших данных

mail.ru group

Криптографический функции, блокчейн

Мацкевич Степан

Алгоритмы и структуры данных

План

- Криптографические хеш-функции
- Базовые алгоритмы защиты
- Блокчейн – основы

Криптографические хеш-функции. MD5

Message Digest

MD1, MD2, MD3, MD4, MD5, MD6 — известные алгоритмы вычисления контрольных сумм.

Один из самых популярных – **MD5**.

128-битный алгоритм хеширования.

Разработан Рональдом Л. Ривестом в 1991г. Использует битовые операции с блоками длины 128.

MD5. Подготовка

1. К сообщению добавляются 000…0 биты и 64-битный размер исходного сообщения так, чтобы размер сообщения был кратен 512 бит. Обозначим его M.
2. Инициализируются 4 переменные по 32 бита A, B, C, D.

```
A = 01 23 45 67; // 67452301h
B = 89 AB CD EF; // EFCDAB89h
C = FE DC BA 98; // 98BADCFEh
D = 76 54 32 10. // 10325476h
```

3. Вводятся 64 числа $K[i] = \text{int}(2^{32} \cdot |\sin n|)$.

MD5. Проход.

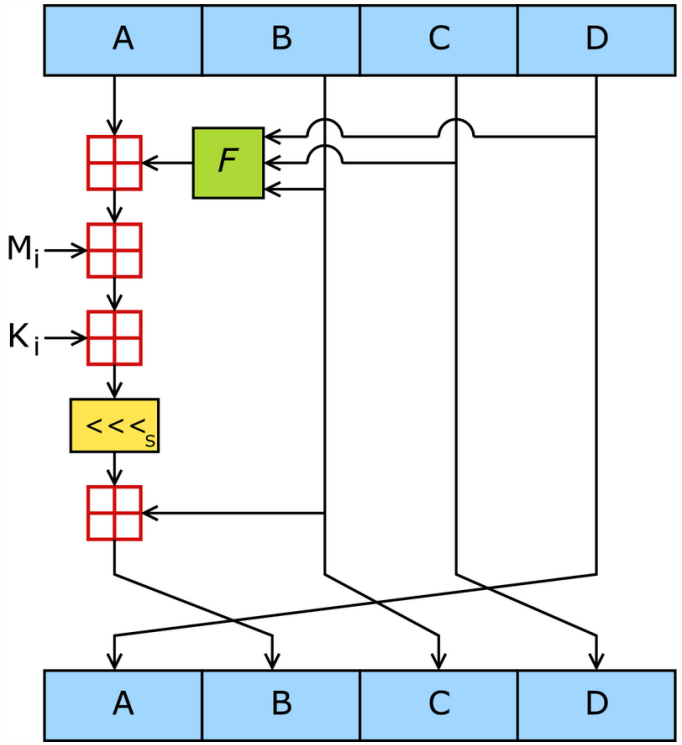
Для каждых 512 бит сообщения выполняется замес.

Обновляем A, B, C, D много раз по циклу, сдвигая переменные по кругу.
4 этапа по 4 раунда (круга). Всего 16 раундов.

Обновление одной переменной

$$p = \big((A + F(B, C, D) + M[k] + K[i]) <<< s\big) + B$$

Здесь $<<<$ – циклический сдвиг.

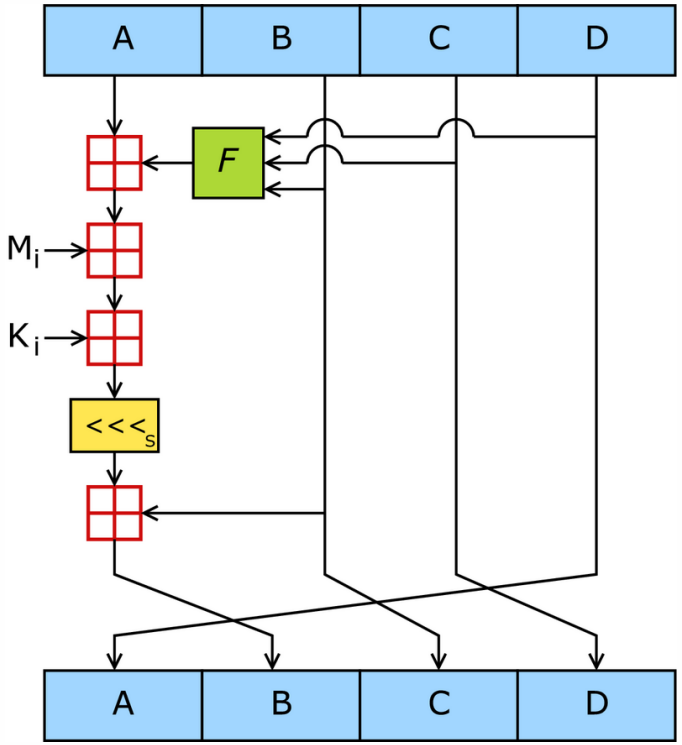


MD5. Проход.

$$p = \big((a + F(b,c,d) + M[k] + K[i]) <<< s\big) + b$$

$F(B,C,D)$ – битовая формула, своя для каждого этапа.

$$F(B,C,D) = (B \wedge C) \vee (\neg B \wedge D)$$
$$G(B,C,D) = (B \wedge D) \vee (C \wedge \neg D)$$
$$H(B,C,D) = B \oplus C \oplus D$$
$$I(B,C,D) = C \oplus (B \vee \neg D)$$



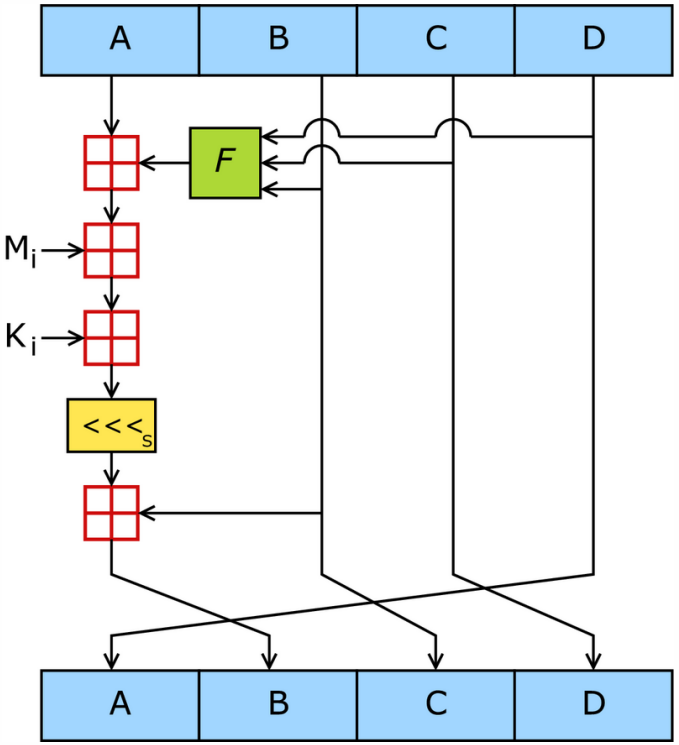
MD5. Проход.

$$p = \big((a + F(b,c,d) + M[k] + K[i]) <<< s\big) + b$$

Параметры k, s, i заданы заранее для каждого раунда.

Для первого этапа:

```
Круг 1: [ 0  7  1][ 1 12  2][ 2 17  3][ 3 22  4]
Круг 2: [ 4  7  5][ 5 12  6][ 6 17  7][ 7 22  8]
Круг 3: [ 8  7  9][ 9 12 10][10 17 11][11 22 12]
Круг 4: [12  7 13][13 12 14][14 17 15][15 22 16]
```



MD5. Окончание.

После выполнения всех 4 раундов значения A, B, C, D используются как начальные для выполнения вычислений со следующими 512 бит сообщения.

Последние значения A, B, C, D являются ответом.

Размер ответа – 128 бит.

Пример для пустой строки:
MD5("") = D41D8CD98F00B204E9800998ECF8427E

MD5

Опубликован в стандарте https://tools.ietf.org/html/rfc1321.

Является улучшением MD4.

Важное преимущество MD5 – лавинный эффект. Замена одного символа приводит к полному изменению хеша:

MD5("md5") = 1BC29B36F623BA82AAF6724FD3B16718

MD5("md4") = C93D3BF7A7C4AFE94B64E30C2CE39F4F

MD5. Использование

MD5 – хорошая «чексумма».

Вероятность ложноположительной ошибки = 2^{-128}.

Использование:

- Проверка корректности переданных/сохраненных данных от СЛУЧАЙНЫХ подмен или ошибок.

Криптоанализ

Криптоанализ хеш-функций направлен на исследование уязвимости для различного вида атак. Основные из них:

- нахождение коллизий — ситуация, когда двум различным исходным сообщениям соответствует одно и то же хеш-значение.
- нахождение прообраза — исходного сообщения — по его хешу.

MD5. Криптографическое использование

Нельзя использовать для защиты информации.

С 1996 года различные исследователи начали успешно строить коллизии.

В 2006 году чешский исследователь Властимил Клима опубликовал алгоритм, позволяющий находить коллизии на обычном компьютере с любым начальным вектором (A,B,C,D).

В 2009 году было показано, что для любых двух заранее выбранных префиксов можно найти специальные суффиксы, с которыми сообщения будут иметь одинаковое значение хеша. Сложность такой атаки составляет всего 2^{39} операций подсчёта MD5.

Stevens M., Lenstra A. K., Weger B. d. Chosen-prefix collisions for MD5 and applications // International Journal of Applied Cryptography — Inderscience Publishers, 2012. — Vol. 2, Iss. 4. — P. 322–359. — ISSN 1753-0563; 1753-0571 — doi:10.1504/IJACT.2012.048084

Криптографические хеш-функции. SHA

Secure Hash Code = SHA

SHA-1, SHA-2, SHA-3 — 160, 256/512, 1024-битные хеши.

Являются более-менее криптографически стойкими к атакам нахождения первообраза и к атаке нахождения коллизий.

SHA-1

Создан и опубликован в 1995.

Авторы – разработчики АНБ+NIST.

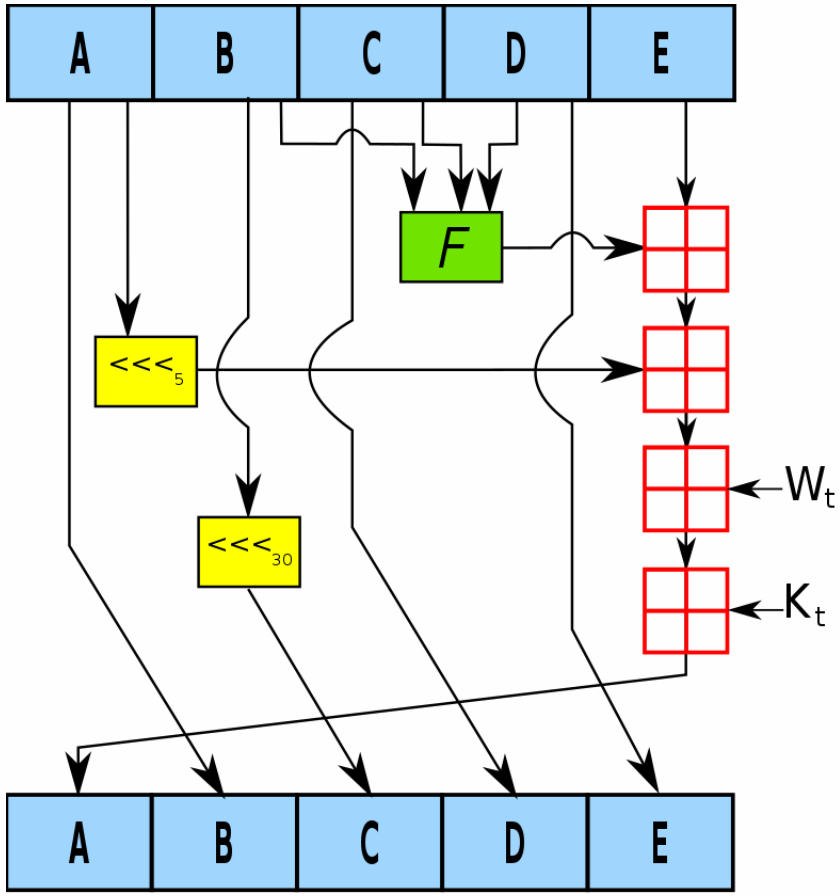
Стандартизован – https://tools.ietf.org/html/rfc3174

Похож на MD5.

Также обрабатываются блоки по 512 бит.

Но количество раундов существенно больше.

4 этапа по 20 раундов.



SHA-1

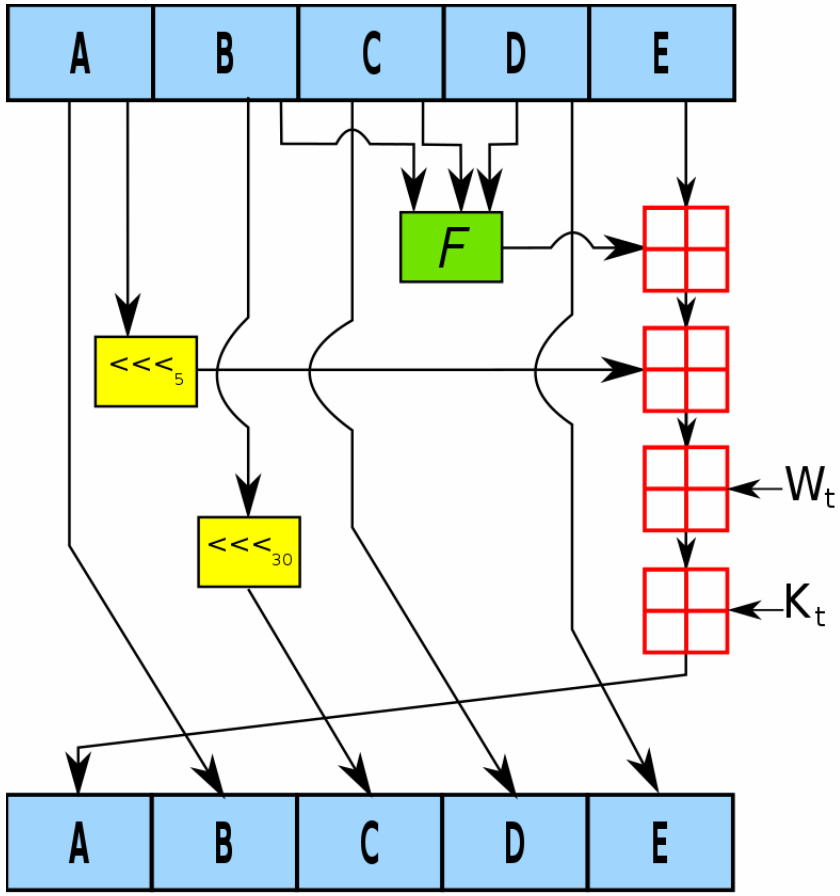
$$p = (A <<< 5) + F_t(B, C, D) + E + W_t + K_t$$

W_t – частично измененное исходное сообщение на старших раундах.

$$W_t = M_t \qquad \text{при } 0 \le t \le 15$$

$$W_t = (W_{t-3} \oplus W_{t-8} \oplus W_{t-14} \oplus W_{t-16}) << 1 \qquad \text{при } 16 \le t \le 79$$

$F_t(m,l,k) = (m \wedge l) \vee (\neg m \wedge k)$	K_t = 0x5A827999	0≤t≤19
$F_t(m,l,k) = m \oplus l \oplus k$	K_t = 0x6ED9EBA1	20≤t≤39
$F_t(m,l,k) = (m \wedge l) \vee (m \wedge k) \vee (l \wedge k)$	K_t = 0x8F1BBCDC	40≤t≤59
$F_t(m,l,k) = m \oplus l \oplus k$	K_t = 0xCA62C1D6	60≤t≤79



Secure Hash Code = SHA

SHA-1 является теоретически взламываемым.

В августе 2005 года на CRYPTO 2005 Xiaoyun Wang, Yiqun Lisa Yin, Hongbo Yu представили версию атаки поиска коллизий на полноценный SHA-1, с вычислительной сложностью в 2^{63} операций. В декабре 2007 года это было проверено Мартином Кохраном.

23 февраля 2017 года специалисты из Google и CWI объявили о практическом взломе алгоритма, опубликовав 2 PDF-файла с одинаковой контрольной суммой SHA-1. Это потребовало перебора $9 \cdot 10^{18}$ вариантов.

Google (группа разработчиков Chrome) объявила в 2014г о постепенном отказе от использования SHA-1.

С 25 апреля 2016 года Яндекс.Почта перестала поддерживать старые почтовые программы, использующие SHA-1.

SHA-2

SHA-2 – семейство однонаправленных хеш-функций, включающее в себя алгоритмы SHA-224, SHA-256, SHA-384, SHA-512, SHA-512/256 и SHA-512/224.

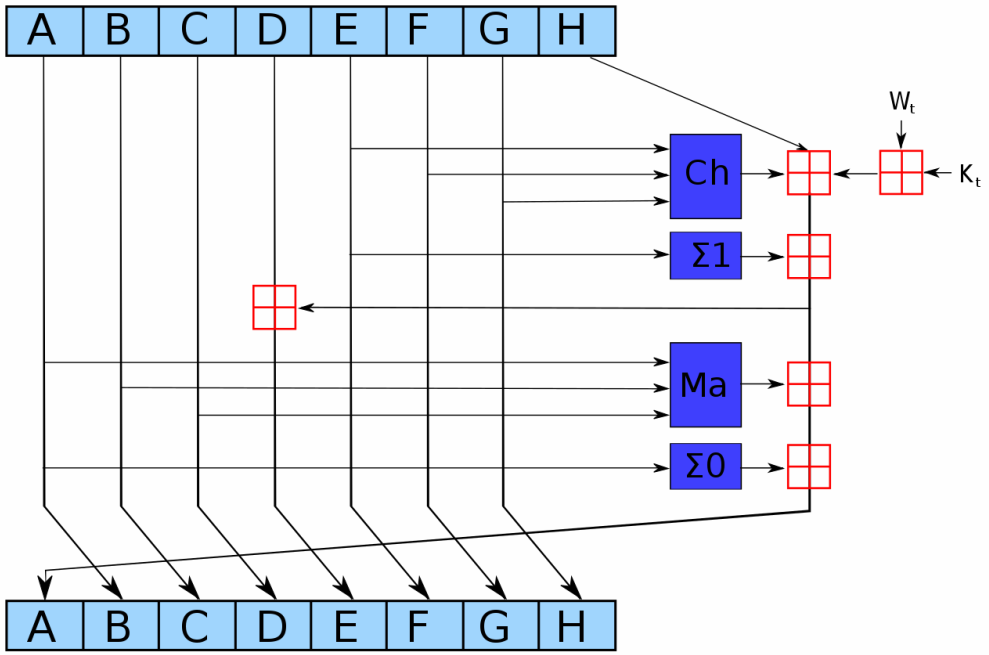
Созданы и опубликованы в 2002-2008гг.

Стандарт – https://tools.ietf.org/html/rfc4634

SHA-2

Также как в MD5 и SHA-1 сообщение разбивается на блоки по 512/1024 бит.

64/80 раундов в обработке одного блока.
Используется больше переменных, больше связей.



SHA-3

Keccak (Кечак) – выбранный NIST по результатам конкурса алгоритм хеширования, ставший SHA-3 (2012г).

Создан и опубликован в 2008г.

Разработчики - Гвидо Бертони, Йоан Даймен, Микаел Питерс, Жиль Ван Аше.

Стандарт – https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.202.pdf

SHA-3

Keccak настраивается следующими параметрами:

1. Размер блока данных,
2. Размер состояния алгоритма,
3. Количество раундов (по умолчанию 24),
4. ...

Размеры хеша аналогичны SHA-2: 224, 256, 384, 512.

Весьма эффективен – быстро вычисляется.

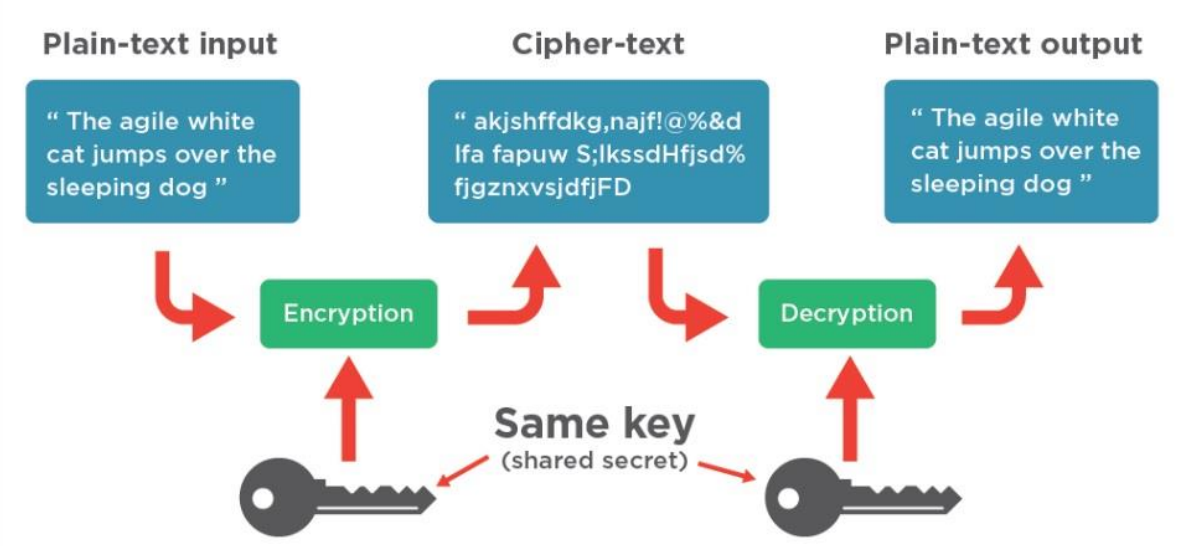
Имеет хорошую криптографическую стойкость. По результатам конкурсов взламывают до 8 раундов.

Базовые алгоритмы защиты

AES – Advanced Encryption Standard

AES – симметричный алгоритм шифрования Rijndal.

- Шифрует блоки по 128 бит.
- Ключ – 128/192/256 бит.
- Принят в качестве стандарта NIST в результате конкурса AES в 2002 году.

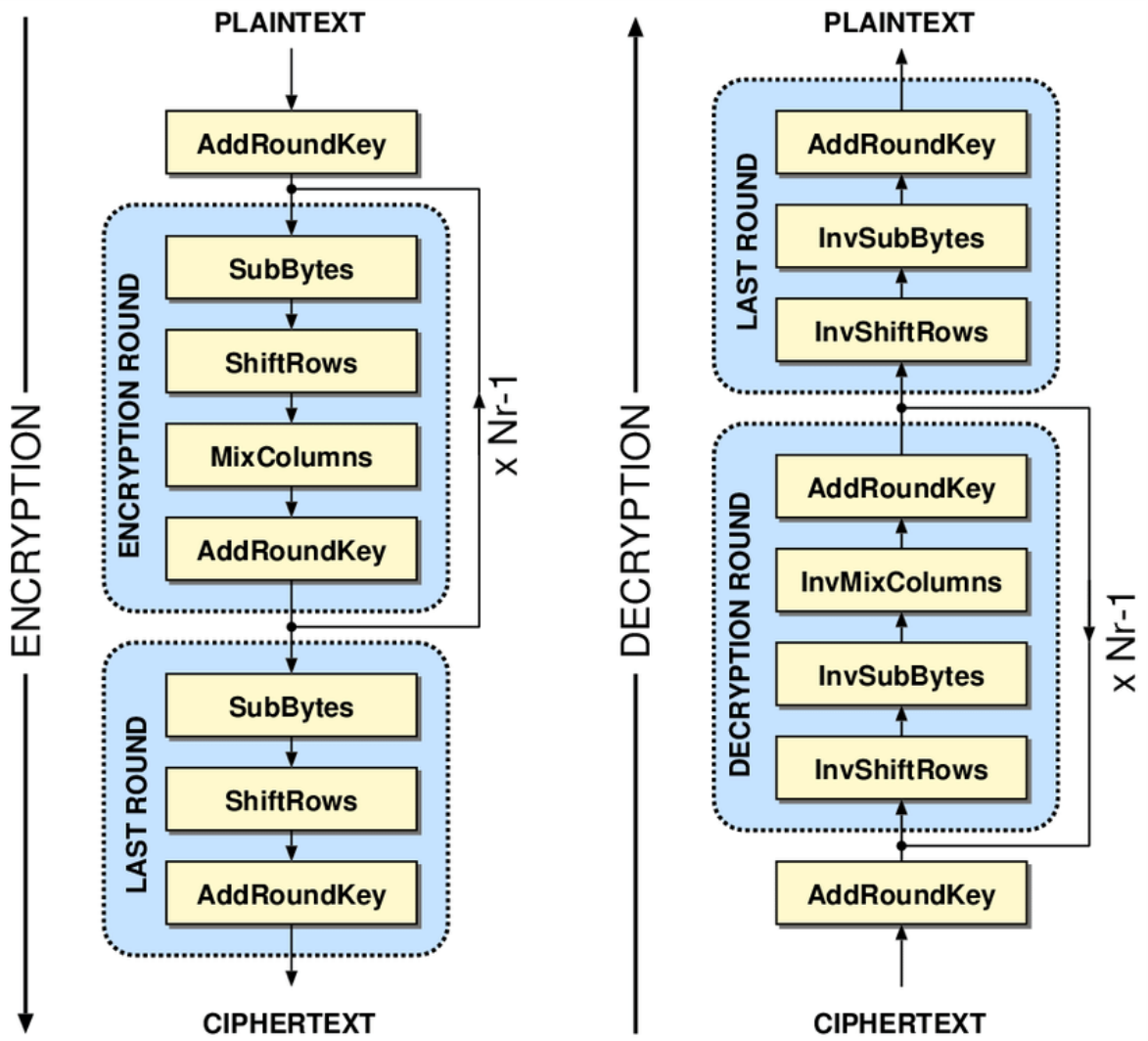


AES – Advanced Encryption Standard

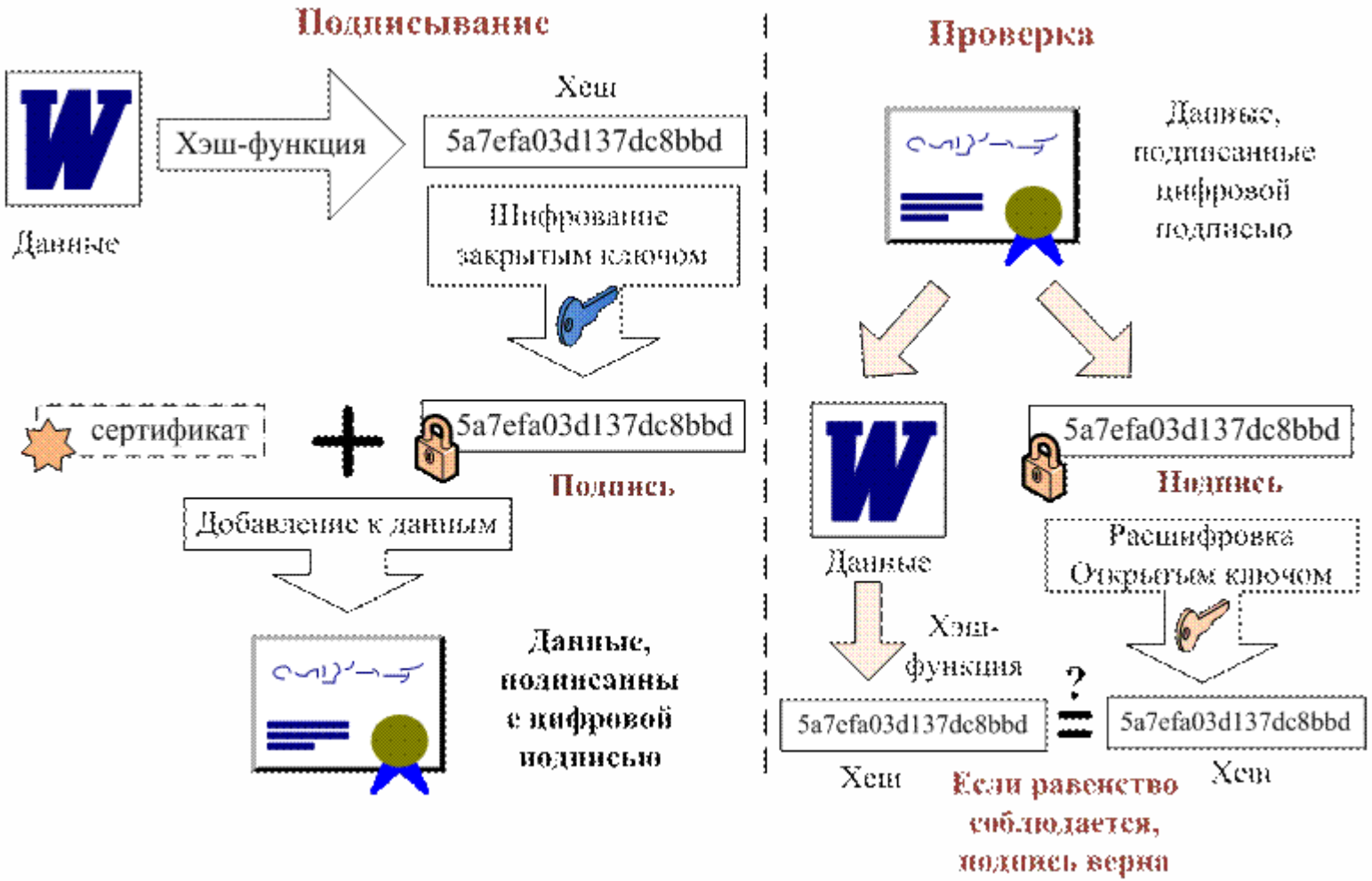
Используются обратимые операции:

- XOR
- Умножение на обратимую матрицу
- В том числе перемешивание колонок

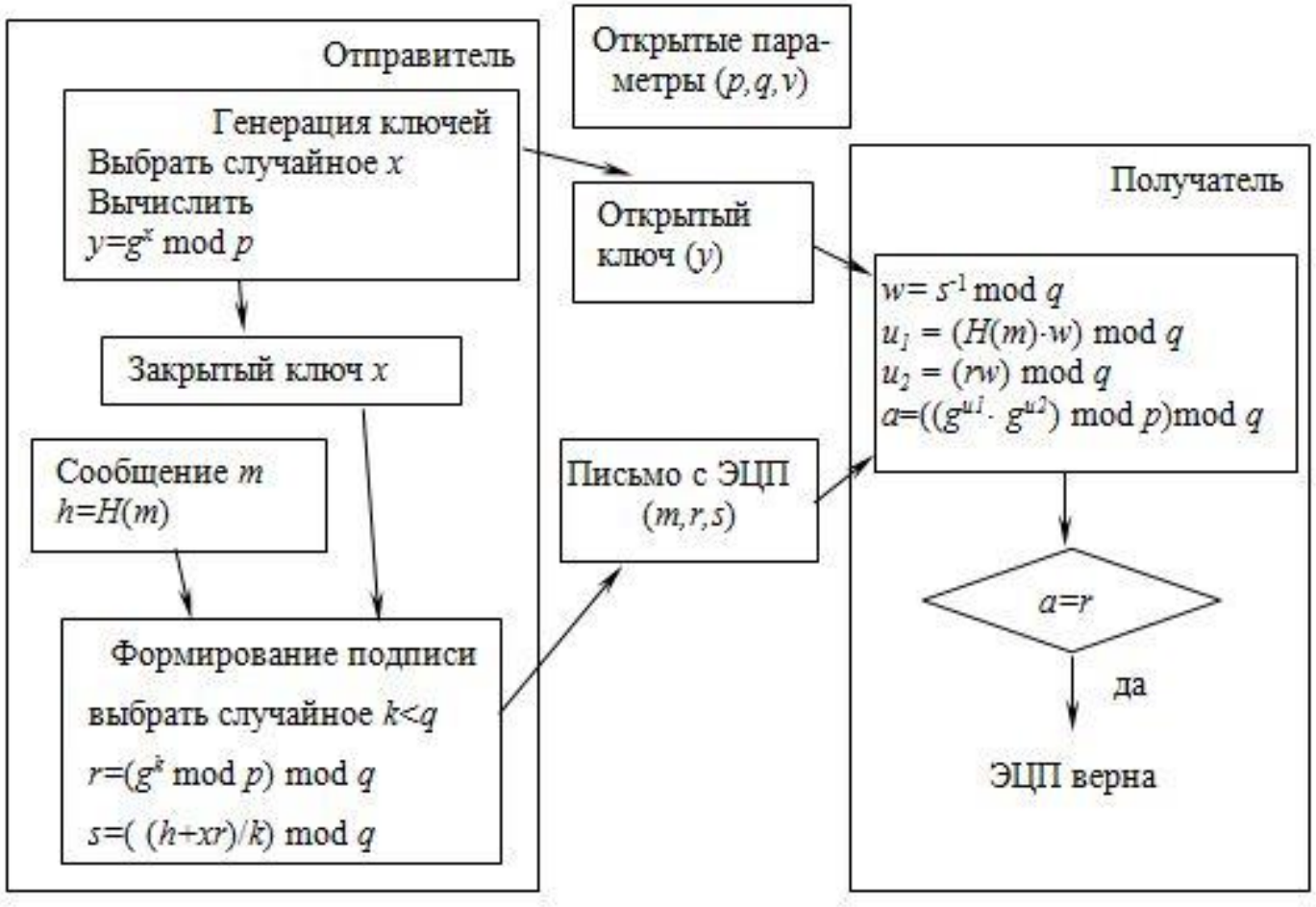
10/12/14 раундов



Цифровая подпись DSA



Цифровая подпись DSA



Цифровая подпись DSA. Подпись.

Подпись сообщения

- Выбор случайного числа k из (0; q), где q – большое простое по размерности h(m)
- Вычисление r = (g^k mod p) mod q, где p просто такое что (p-1) делится на q, а g такое, что его мультипликативный порядок по модулю p равен q
- Вычисление s = (k^{-1}(h(m) + x r)) mod q, где x – закрытый ключ
- Выбор другого k, если оказалось, что r = 0 или s = 0

Подписью является пара чисел (r, s)

Цифровая подпись DSA. Проверка.

Проверка подписи

- Вычисление w = s^{-1} mod q
- Вычисление u_1 = (h(m) w) mod q
- Вычисление u_2 = (r w) mod q
- Вычисление v = ((g^{u_1} y^{u_2}) mod p) mod q, где y = g^x mod p – открытый ключ

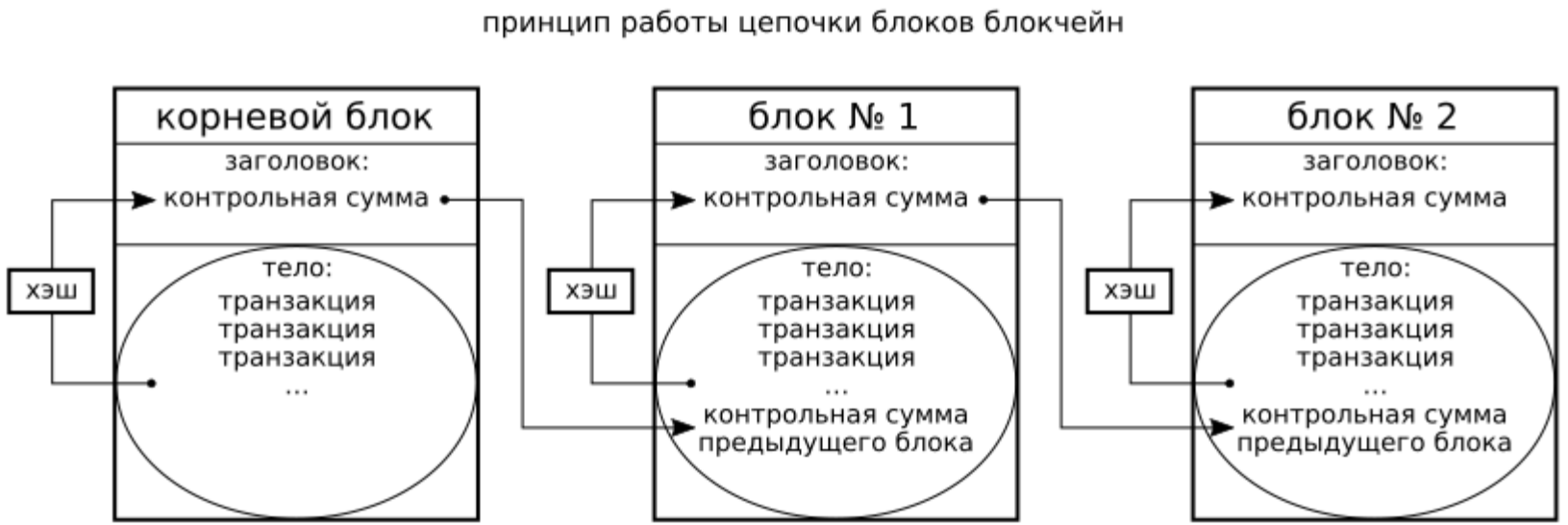
Подпись верна, если v = r

Открытыми параметрами являются числа (p, q, g, y). Закрытый параметр только один — число x. При этом числа (p, q, g) могут быть общими для группы пользователей, а числа x и y являются соответственно закрытым и открытым ключами конкретного пользователя.

Блокчейн

Блокчейна

Блокчейн – цепочка криптографически связанных блоков.



Правила добавления блока

Пример современного хеша блока BTC –
00000000000000000000526273c1abbdb82cc3f7964b5c287193eeaf0f86d14b3

Хешируются с помощью SHA-256 данные:

- Список добавленных транзакций до 1Мб
- Хеш предыдущего блока
- Timestamp
- Соль (подбирается)

Если значение меньше порогового значения, то блок может быть добавлен в цепочку.

Порог = Максимум / Сложность.

Вычисляется на основе истории так, чтобы очередные блоки находились в среднем раз в 10 минут.

Блокчейн биткойна (BTC)

В BTC по состоянию на 16.11.19:

- Находится 604 048 блоков
- Средний размер блока 2500 транзакций
- Общий размер блокчейна примерно 300 Гб.
- В сутки генерируется примерно 170 блоков (~1 блок в 8.5 минут)

Заглянем в текущие блоки

https://blockexplorer.com/blocks

Спасибо за внимание!