Направление подготовки бакалавров
090301 «Информатика и вычислительная техника»

ХЭШ-функции

Караваева Ольга Владимировна

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Основные понятия и определения

Хэширование – преобразование массива входных данных произвольной длины в (выходную) битовую строку фиксированной длины, выполняемое определённым алгоритмом.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Основные понятия и определения

- **Хэш-функция –** функция, преобразующая по детерминированному алгоритму входной массива данных определенной длины (ключ) в выходную битовую строку фиксированной длины(значение).

- **Коллизия –** это ситуация, когда разным ключам соответствует одно значение хэш-функции

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Основные понятия и определения

Понятие «хорошей» хэш-функции

- функция должна быть простой с вычислительной точки зрения;
- функция должна распределять ключи в хеш-таблице наиболее равномерно;
- функция должна минимизировать число коллизий – то есть ситуаций, когда разным ключам соответствует одно значение хэш-функции.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Хэш-функции
Метод Деления с остатком

При построение хэш-функции методом деления с остатком ключу k ставится в соответствие остаток от деления k на m, где m — число возможных значений хэш-функции:

$h(k)=k \bmod m$

Например, при размере хэш-таблицы $m=12$ и ключе $k=50$

$h(k)=2$.

Некоторых значений основания m следует избегать. Например, если $m=2^p$, то хэш-функция — это просто p младших битов ключа k. Хорошие результаты дает выбор в качестве m простого числа, далекого от степеней двойки.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Хэш-функции
Метод умножения

Пусть количество хэш-значений равно m.

Зафиксируем константу A в интервале 0< A < 1

$h(k) = [m(kA \ mod \ 1)]$, где (k A mod 1) - дробная часть kA

Достоинство метода умножения в том, что качество хэш-функции мало зависит от выбора m. Обычно в качестве m выбирают степень двойки, поскольку в большинстве компьютеров умножение на такое m реализуется как сдвиг слова.

Кнут предложил в качестве A использовать, число A = 0,6180339887…

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Хэш-Функции универсальное хэширование

Основная идея универсального хэширования — выбирать хэш-функцию во время исполнения программы случайным образом из некоторого множества.

При повторном вызове с теми же входными данными алгоритм будет работать уже по-другому.

При случайном выборе хэш-функции вероятность коллизии между двумя данными ключами должна равняться вероятности совпадения двух случайно выбранных хэш-значении (которая равна 1/m).

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Хэш-таблица

Хэш - таблица – это структура данных, хранящая ключи в таблице. Индекс ключа вычисляется с помощью хэш-функции. Операции: добавления, удаление, поиск.

Пусть хэш-таблица имеет размер M, а количество элементов в хэш-таблице - N.

Коэффициент заполнения хэш-таблицы – это количество хранимых элементов массива, деленное на число возможных значений хеш-функции. Обозначим его $\alpha = n/m$.

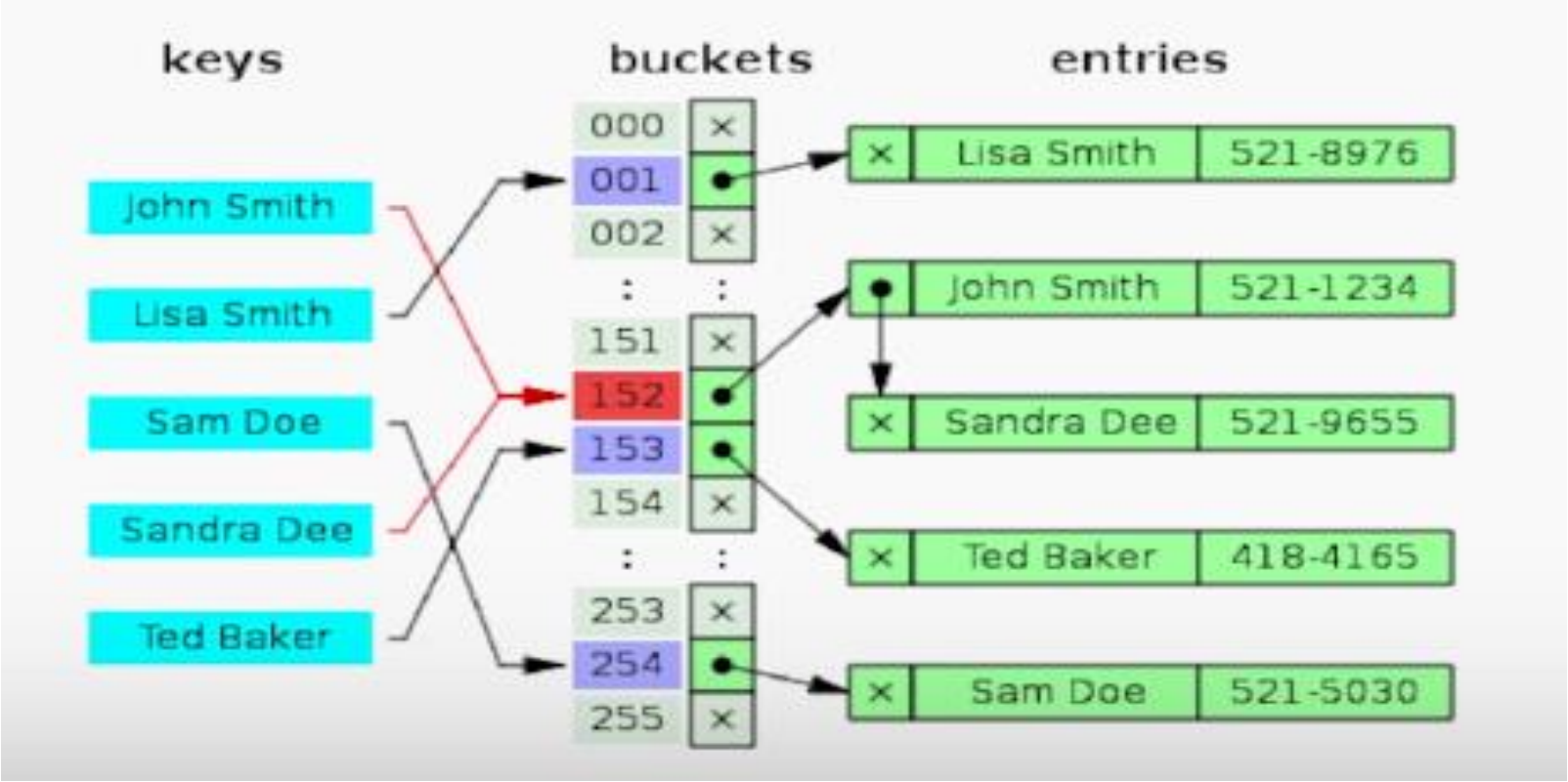
ВЯТСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

Методы разрешения коллизий

- Метод цепочек
- Метод открытой адресации

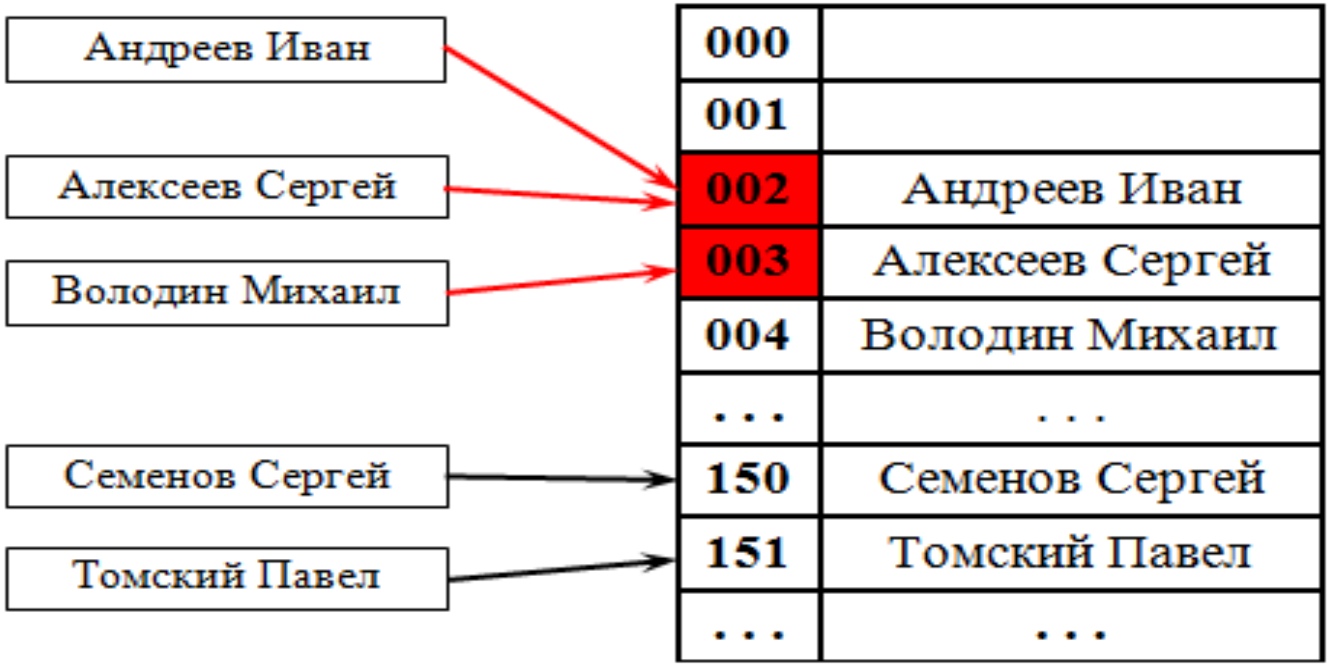
ВЯТСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

Метод цепочек



ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Метод открытой адресации



ВЯТСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

Области применения

- Криптография
- Хранение паролей
- В системах передачи данных для контроля целостности
- Системы электронной подписи

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Криптография

Хэш-функция h(x) называется криптографической, если она удовлетворяет следующим требованиям:

• необратимость: для заданного значения хэш-функции c должно быть сложно опре- делить такой ключ x, для которого h(x) = c

• стойкость к коллизиям первого рода: для заданного ключа x должно быть вычис- лительно невозможно подобрать другой ключ y, для которого h(x) = h(y);

• стойкость к коллизиям второго рода: должно быть вычислительно невозможно по- добрать пару ключей x и y, имеющих одинаковый хэш.

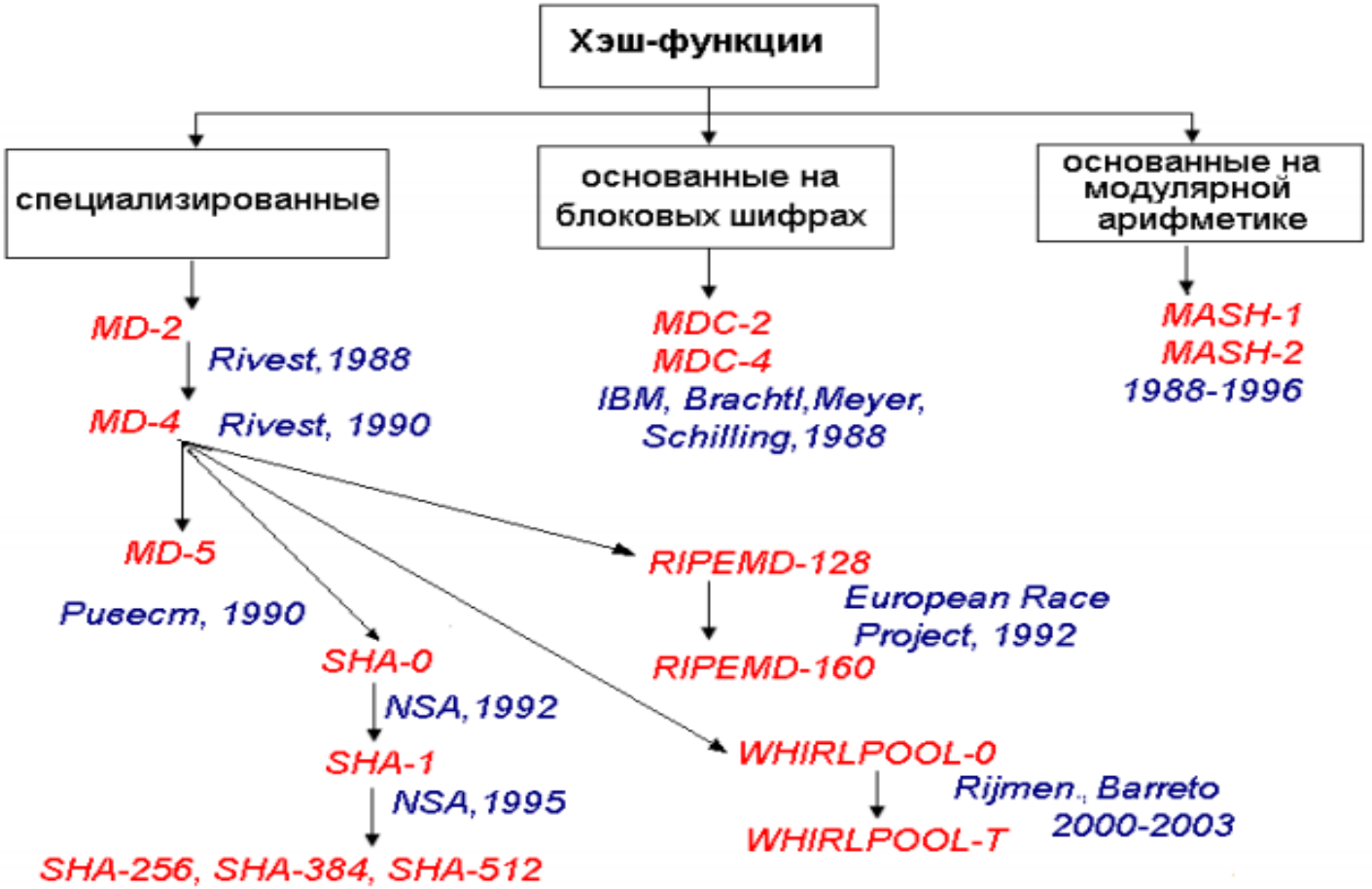
Криптография

Криптографические хэш-функции обычно не используются в хэш-таблицах, потому что они сравнительно медленно вычисляются и имеют большое множество значений.

Типы криптографических хэш-функций

Ключевые хеш-функции
или
коды аутентификации сообщений
(***Message Authentication Code - MAC***)
Используются в системах с доверяющими друг другу пользователями

Бесключевые хеш-функции
или
коды обнаружения ошибок
(***Modification Code – MDC***
или
Message Integrity Code –MIC)
Используются в системах как с доверяющими друг другу, так и с недоверяющими пользователями


Хэш-функции
специализированные
основанные на блоковых шифрах
основанные на модулярной арифметике
MD-2
Rivest, 1988
MD-4
Rivest, 1990
MD-5
Ривест, 1990
SHA-0
NSA, 1992
SHA-1
NSA, 1995
SHA-256, SHA-384, SHA-512
MDC-2
MDC-4
IBM, Brachtl, Meyer, Schilling, 1988
MASH-1
MASH-2
1988-1996
RIPEMD-128
European Race Project, 1992
RIPEMD-160
WHIRLPOOL-0
Rijmen., Barreto 2000-2003
WHIRLPOOL-T

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Алгоритм SHA-1

Основные характеристики алгоритма:

Длина хэш-кода – 160 бит

Длина обрабатываемых блоков – 512 бит

Число шагов алгоритма - 80 (4 раунда по 20 шагов)

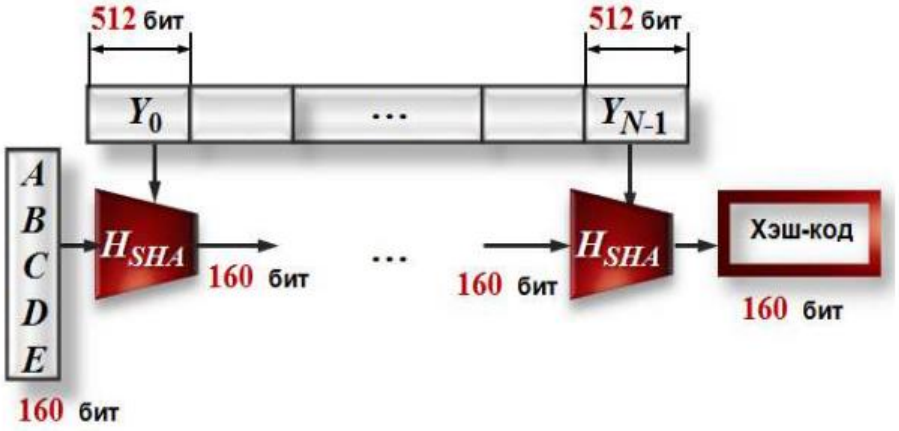
Максимальная длина хэшируемых данных – $2^{64} - 1$.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Логика выполнения SHA-1

Алгоритм получает на входе сообщение максимальной длины $2^{64} - 1$– бит и создает в качестве выхода дайджест сообщения длиной 160 бит.

Алгоритм состоит из следующих шагов:



ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 1: добавление недостающих битов

На вход алгоритма хеширования SHA-1 подается сообщение длиной 2590 битов.

$$m + s + 64 \equiv 0(mod512) => s$$
$$= -m - 64(mod512) = -2590 - 64(mod512)$$
$$= 418$$

Дополнение состоит из одной 1 и 417 нулей.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 1: добавление недостающих битов

Сообщение добавляется таким образом, чтобы его длина была кратна 448 по модулю 512
Это означает, что длина добавленного сообщения на 64 бита меньше, чем число, кратное 512

Добавление осуществляется всегда, даже если сообщение уже имеет нужную длину. Таким образом, число добавляемых битов находится в диапазоне от 1 до 512.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 1: добавление недостающих битов

На вход алгоритма хэширования SHA-1 подается сообщение длиной 2590 битов.

$$m + s + 64 \equiv 0(mod512) => s = -m - 64(mod512)$$
$$= -2590 - 64(mod512) = 418$$

Дополнение состоит из одной 1 и 417 нулей.

Шаг 2: добавление длины

К сообщению добавляется блок из 64 битов. Этот блок трактуется как беззнаковое 64-битное целое и содержит длину исходного сообщения до добавления.

Результатом первых двух шагов является сообщение, длина которого кратна 512 битам.

Расширенное сообщение может быть представлено как последовательность 512-битных блоков $Y_0, Y_1, \ldots, Y_{L-1}$, так что общая длина расширенного сообщения есть L * 512 бит. Таким образом, результат кратен шестнадцати 32-битным словам.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 3: инициализация SHA-1 буфера

В алгоритме используется 160-битный буфер для хранения промежуточных и окончательных результатов хэш-функции. Буфер может быть представлен как пять 32-битных регистров A B C D E

A = 67 45 23 01;

B= EF CD AB 89;

C= 98 BA DC FE ;

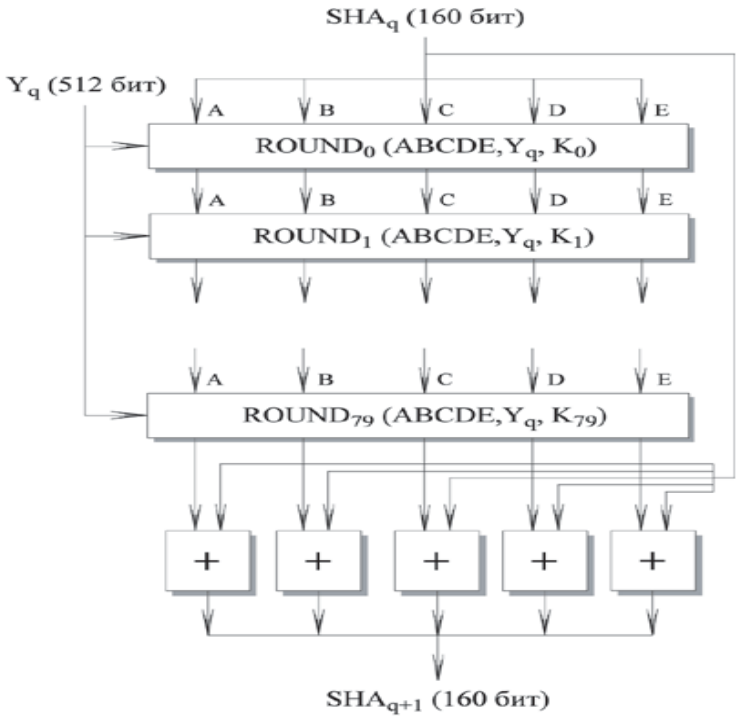
D =10 32 54 76 ;

E = C3 D2 E1 F0.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 4: обработка сообщения в 512-битных (16-словных) блоках

Основой алгоритма является модуль, состоящий из 80 циклических обработок, обозначенный как H_{SHA}. Все 80 циклических обработок имеют одинаковую структуру.



ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 4: обработка сообщения в 512-битных (16-словных) блоках

Каждый цикл получает на входе текущий 512-битный обрабатываемый блок Yq и 160-битное значение буфера ABCDE, и изменяет содержимое этого буфера.

В каждом цикле используется дополнительная константа Kt, которая принимает только четыре различных значения:

0 <= t <= 19 Kt = 5A827999

(целая часть числа [2^30 x 2^1/2])

20 <= t <= 39 Kt = 6ED9EBA1

(целая часть числа [2^30 x 3^1/2])

40 <= t <= 59 Kt = 8F1BBCDC

(целая часть числа [2^30 x 5^1/2])

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 4: обработка сообщения в 512-битных (16-словных) блоках

60 <= t <= 79 Kt = CA62C1D6
(целая часть числа [2^30 x 10^1/2])

Для получения SHAq+1 выход 80-го цикла складывается со значением SHAq. Сложение по модулю 2^32 выполняется независимо для каждого из пяти слов в буфере с каждым из соответствующих слов в SHAq.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Шаг 5: выход

После обработки всех 512-битных блоков выходом L-ой стадии является 160-битный *дайджест сообщения*.

Дайджест сообщения — это уникальная последовательность символов, однозначно соответствующая содержанию сообщения. Обычно дайджест имеет фиксированный размер, который не зависит от длины самого сообщения.

Дайджест вставляется в состав ЭЦП вместе со сведениями об авторе и шифруется вместе с ними.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Сравнение SHA-1 и MD5

- **Безопасность**: наиболее очевидное и наиболее важное различие состоит в том, что *дайджест SHA-1* на 32 бита длиннее, чем *дайджестMD5*. Если предположить, что оба алгоритма не содержат каких-либо структурированных данных, которые уязвимы для криптоаналитических атак, то *SHA-1* является более стойким алгоритмом. Используя лобовую атаку, труднее создать произвольное сообщение, имеющее данный *дайджест*, если требуется порядка 2^{160} операций, как в случае алгоритма *SHA-1*, чем порядка 2^{128}операций, как в случае алгоритма *MD5*. Используя лобовую атаку, труднее создать два сообщения, имеющие одинаковый *дайджест*, если требуется порядка 2^{80} как в случае алгоритма *SHA-1*, чем порядка 2^{64} операций как в случае алгоритма *MD5*.
- **Скорость**: так как оба алгоритма выполняют сложение по модулю 2^{32}, они рассчитаны на 32-битную архитектуру. *SHA-1* содержит больше шагов (80 вместо 64) и выполняется на 160-битном буфере по сравнению со 128-битным буфером *MD5*. Таким образом, *SHA-1*должен выполняться приблизительно на 25% медленнее, чем *MD5* на той же аппаратуре.
- **Простота и компактность**: оба алгоритма просты и в описании, и в реализации, не требуют больших программ или подстановочных таблиц. Тем не менее, *SHA-1* применяет одношаговую структуру по сравнению с четырьмя структурами, используемыми в *MD5*. Более того, обработка слов в буфере одинаковая для всех шагов *SHA-1*, в то время как в *MD5* структура слов специфична для каждого шага.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Примеры

Российский стандарт ГОСТ Р 34.11-94 определяет алгоритм и процедуру вычисления хэш-функции для любых последовательностей двоичных символов, применяемых в криптографических методах обработки и защиты информации. Этот стандарт базируется на блочном алгоритме шифрования ГОСТ 28147-89.

Данная хэш-функция формирует 256-битовое хэш-значение. Данная хэш-функция определена стандартом ГОСТ Р 34.11-94 для использования совместно с российским стандартом электронной цифровой подписи.

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Примеры

ГОСТ Р 34.11-2012 – обновлённая версия, отличающаяся высокой стойкостью к попыткам взлома и стабильностью в работе. Объем выдаваемого хеша может быть как 512, так и 256 бит.

ВЯТСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

Дополнительные источники

https://moodle.kstu.ru/mod/page/view.php?id=34862