Московский Государственный Университет
имени М. В. Ломоносова

Механико-математический факультет
Кафедра математических и компьютерных методов анализа

Дипломная работа

студента 601 группы
Медведева Егора Михайловича

Сравнительный анализ хеш-функций
Comparative analysis of hash functions

Научный руководитель
профессор, д.ф.-м.н. Чубариков В. Н.
Рецензент
ФИО РЕЦЕНЗЕНТА И ТД

Москва, 2018 год.

Содержание

1 Введение

2 Постановка задачи и формулировка основных результатов

Применение.

3 Основные и вспомогательные определения

В данной главе введем основные определения, которые будут использоваться в этой работе. Начнем с формального определения хеш-функции:

Пусть
$\{0,1\}^n$ - множество строк длины m, состоящих из битов 0 или 1.
$\{0,1\}^*$ - множество всех строк конечной длины, состоящих из битов 0 или 1.

Определение 1. Криптографической хеш-функцией h называется преобразование вида

$$h : \{0,1\}^* \to \{0,1\}^n$$

обладающее следующими свойствами:

1. Значение $h(M)$(где $M \in \{0,1\}^*$) должно "вычисляться легко"
2. При изменении всего лишь одного бита входного сообщения значение $h(M)$ меняет хотя бы половину своих битов
3. Прообраз для заданного $h(M)$ должен "вычисляться сложно" (pre-image resistance)
4. Второй прообраз $M' \neq M$, такой что $h(M') = h(M)$, должен "вычисляться сложно" (second pre-image resistance)
5. Нахождение M и $M' \neq M$, таких что $h(M) = h(M')$, "вычисляется сложно" (collision resistance)

Замечание 1. Далее под "хеш-функцией" для удобства имеем в виду "криптографическую хеш-функцию".

Входную строку M будем называть "сообщением". Значение хеш-функции $h(M)$ будем называть "хешем", "хеш-кодом" или "хеш-суммой".

Под фразами "вычисляется сложно" или "вычислительно неразрешима" далее подразумеваем, что задача не решается за разумное время на современной вычислительной технике (очевидно, что задачи из свойств 2), 3) и 4) можно решить, например, полным перебором).

Пример 1. Пусть h - хеш-функция, определенная российским стандартом ГОСТ Р 34.11-2012 256 (ее описание см. далее). Ниже представлены результаты хеширования некоторых строк (результат представлен в шестнадцатеричной системе счисления):

1. $M =$"Московский Государственный Университет имени М. В. Ломоносова" $\Rightarrow$
$h(M) =$"2609a10022385596400318f6b959b9d449edbf7820ec188c7d8ddbc06a09ab0b"

2. $M=$ "МГУ им. М. В. Ломоносова" $\Rightarrow$
   $h(M)=$ "15414d11b2cbd98c858870463ed42189023845521f1bcb914817897c9f312d43"

3. $M=$ "МГУ им М. В. Ломоносова" $\Rightarrow$
   $h(M)=$ "4b54a14ab2320e4ed25e542410e424aad19ccc04fcd4debf46430efa6d972326"

Замечание 2. Отличие двух последних примеров состоит в присутствии и отсутствии знака точки: "им." и "им". Но, как следует из свойства 2) определения 1, хеши различаются очень сильно.

Следующее определение играет важную роль при построении хеш-функций.

Определение 2. Блочным шифром называются алгоритмы шифрования и расшифрования с одинаковым ключом $K \in \{0,1\}^m$ (то есть так называемый симметричный шифр), представляемые в виде функций E_K и D_K и являющимися для каждого фиксированного ключа биективным отображением на множестве $\{0,1\}^n$:

$$E_K(M) := E(K,M) : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n,$$
$$D_K(C) := D(K,C) : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n,$$

причем $D = E^{-1}$, то есть $\forall K:$

$$D_K(E_K(M)) = M \text{ и}$$
$$E_K(D_K(C)) = C.$$

4 Атака дней рождений

5 Принципы построения хеш-функций

В общем случае в основе построения хеш-функций лежит итеративная последовательная схема, когда на вход каждой итерации поступает блок исходного текста и результат предыдущей итерации. Ядром каждой итерации служит функция сжатия f, принимающая на вход блок определенной длины m и результат предыдущей итерации длины n, то есть:

$$f : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n$$

Данная конструкция называется Структурой Меркла-Дамгарда, которая была придумана независимо Ральфом Мерклом и Иваном Дамгардом. Ими же было установлено, что если функция сжатия устойчива к коллизиям, то и хеш-функция будет также устойчива к коллизиям.

Опишем подробнее все шаги схемы:

Пусть $M \in \{0,1\}^*$ - исходное сообщение для хеширования.

1. Разобьем сообщение M на блоки $M_1, \ldots, M_s$ длины m.
2. Дополним входное сообщение заранее определенным образом (например, нулями), если длина M не кратна m.
3. Определить начальное значение H_0
4. i-й шаг итерации ($i = 1, \ldots, s$) заключается в вычислении значения $H_i = f(M_i, H_{i-1})$
5. Значение H_s (возможно, после некоторых дополнительных преобразований) и является конечным хешем сообщения M.

Определение 3. Структурой Меркла-Дамгарда называется приведенный выше алгоритм вычисления хеша (рис. 1).

картинка
картинка
картинка
картинка
картинка
картинка
картинка
картинка

Также существуют улучшенные схемы, основанные на структуре Меркла-Дамгарда:

- Структура Девиса-Мейера
- Структура Матиса-Мейера-Осеаса
- Структура Миагучи-Пренеля

Во всех этих схемах в качестве функции сжатия f используется блочный шифр E. Например, могут использоваться следующие популярные стандарты шифрования - DES, AES или ГОСТ 28147-89.
Ниже приведем более подробное описание каждой из схем.

5.1 Структура Девиса-Мейера

Как сказано выше, данная структура использует блочный шифр E. В качестве ключа шифр использует входной блок, а на вход подается результат предыдущей итерации (для первой итерации - это некоторое начальное значение). Затем для полученного результата выполняем операцию побитового сложения с значением предыдущей итерации (см. рис 2). Результат сложения и будет финальным значением итерации. Математически все это записывает так:

$$H_i = E_{M_i}(H_{i-1}) \oplus H_{i-1}$$

картинка
картинка
картинка
картинка
картинка

5.2 Структура Матиса-Мейера-Осеаса

Отличие этой структуры от предыдущей заключается в том, что входной блок и значение предыдущей итерации меняются местами. Однако получается несоответствие длин, так как ключ и выходное значение шифра имеют разные длины. В связи с этим надо иметь какое-то дополнительное преобразование g:

$$g : \{0,1\}^n \to \{0,1\}^m$$

Тогда перед выполнением шифра E применяем функцию g к результату предыдущей итерации и уже полученное значение используем в качестве ключа шифра.
Далее к результату шифра применяем операцию побитового сложения с входным блоком (см. рис. 3). Математически данные преобразования выглядят так:

$$H_i = E_{g(H_{i-1})}(M_i) \oplus M_i$$

картинка
картинка
картинка
картинка
картинка

5.3 Структура Миагучи-Пренеля

Эта структура считается самой популярной и надежной из представленных здесь схем. Она аналогична предыдущей структуре за исключением того, что во время побитового сложения добавляется еще одно слагаемое - результат предыдущей итерации. Таким образом, математически весь алгоритм записывается так:

$$H_i = E_{g(H_{i-1})}(M_i) \oplus M_i \oplus H_{i-1}$$

картинка
картинка
картинка
картинка
картинка