ХЭШИРОВАНИЕ И ХЭШ-ФУНКЦИИ

**Цель работы**

Изучить базовые принципы хеширования и простейшие хэш-функции.

**Теоретическая часть**

Хеширование – преобразование входных данных произвольной длинны в битовою строку фиксированной длинны.

Хэш функция – функция, которая принимает входные данные произвольной длинны и возвращает битовою строку фиксированной длинны.

Битовая строка – последовательный набор битов (в простейшем случае – число или символ).

Хэш функции применяются для решения следующих задач:

1. Контрольные суммы;
2. Криптографические алгоритмы;
3. Поиск и сравнение данных.

**Простейшие примеры**

1. Рассмотрим следующую задачу: есть файл с данными, который передаётся посредством сети интернет. После получения этого файла, необходимо проверить, был ли файл повреждён при передаче.

Один из способов проверки: рассматриваем файл, как последовательность бит, разбиваем на блоки по 32 бита и складываем, получившиеся числа, отбрасывая переполнение.

Если файл начинается со следующих бит:

10100111 01011100 10010100 01111111 10010001 10001101 10010100 00010000 00011010 00001000 10010100 01111111 00010001 10001101 10010100 01111111

Берём первые 32 бита:

10100111 01011100 10010100 01111111

И прибавляем к ним следующие 32 бита:

10010001 10001101 10010100 00010000

Получается:

1 00111000 11101010 00101000 10001111

Отбрасываем переполнение:

00111000 11101010 00101000 10001111

Таким образом можно просуммировать весь файл. В итоге у нас получится число (последовательность бит). Это число называется хэш-сумма. Его нужно записать в конец файла. При получении файла на другой машине, необходимо заново посчитать хэш сумму и сравнить результат.

1. Рассмотрим другую задачу: необходимо сохранить пароль на сервере. Если хранить пароль текстовой строкой, то он может быть украден. Поэтому используют специальную хэш-функцию, которая строчку пароля произвольной длинны преобразует в число, которое можно хранить без угрозы хищения. (см. подробнее «Алгоритм md5»).
2. Рассмотрим последнюю задачу: есть строчки произвольной длинны, необходимо найти среди них одинаковые. Необходимо помнить, что для сравнения строк, необходимо сравнить последовательно каждый символ, а это занимает много времени.

Для решения этой задачи используют следующий приём: для каждой строки считают хэш-функцию (сумма кодов символов по модулю простого числа) и сравнивают значения этих хэш-функций. При равенстве значений хэш функций, сравнивают сами строки.

**Контрольные вопросы**

1. Что такое хэш-функция?
2. Что такое битовая строка?
3. Для чего применяется хэш-функция?
4. Как сравнивают строки произвольной длинны с использование хэш-функций?
5. \*Почему при построении хэш-функции строк используют модуль простого числа?