Хеширование строки. Хеш функции

Олексій Лубинець, 5 курс, ФВЕ

Об'єктно-орієнтоване програмування

28 березня 2018 р.

Визначення

Хеширование - преобразование массива входных данных произвольной длины в строку установленной длины, выполняемое определённым алгоритмом.
Функция, воплощающая алгоритм и выполняющая преобразование, называется **хеш-функцией** или **функцией свёртки**.
Исходные данные называются **входным массивом**, **ключом** или **сообщением**. Результат преобразования (выходные данные) называется **хешем**, **хеш-кодом**, **хеш-суммой**, **сводкой сообщения**.

Визначення

Мы можем подать на вход 128-битной хеш-функции роман Льва Толстого в шестнадцатеричном виде или число 1.
В результате на выходе мы в обоих случаях получим разные наборы псевдослучайных шестнадцатеричных цифр вида: c4ca4238a0b923820dcc509a6f75849b.

Властивості хеш-функції (вимоги)

Неустойчивость

При изменении исходного текста даже на один знак, полностью меняется результат хеш-функции

Односторонняя функция

Восстановление по хешу входного массива - очень сложная задача (в плане вычислительной сложности)

Устойчивость к коллизиям первого рода

По заданному x должно быть трудно найти z со свойством f(x) = f(z)

Устойчивость к коллизиям второго рода

Должно быть трудно найти произвольные x = z, такие что f(x) = f(z)

Приклади хеш-функцій

1) $hash(x) = x \ mod \ P$
2) $hash(*s) = (s[0] + s[last]) \ mod \ P$
3) $hash(*s) = (s[0] + s[1] + ...) \ mod \ P$

(1-3) - “плохие хеш-функции”

4) $hash(*s) = (s[0] + s[1]P + s[2]P^2 + ... + s[n]P^n) \ mod \ Q$

Застосування

- построение ассоциативных массивов (где индекс - не только целое число, но и, например, строка)
- поиск дубликатов в сериях наборов данных
- построение уникальных идентификаторов для наборов данных
- вычисление контрольных сумм от данных (сигнала) для последующего обнаружения в них ошибок (штрих-код, ИНН, номер карты etc)
- сохранении паролей в системах защиты в виде хеш-кода
- выработка электронной подписи.

Колізії хеш-функцій

Коллизия хеш-функции — два различных входных блока данных x и y для хеш-функции H таких, что H(x)=H(y). Идеальное хеширование - хеширование без коллизий (мат. инъективное отображение).

Проблемы из-за коллизий

- подделка сообщений
- подделка ЭЦП, сертификатов
- взлом паролей

Поиск коллизий

- атака грубой силы (brute force) $\mathcal{O}(2^N)$
- атака дней рождения $\mathcal{O}(2^{N/2})$
- метод удлинения сообщения