Filtros de Bloom y Cryptohashing

Omar López Rubio, Marc Ángel Ortiz, David Williams

Abril 2017

1 Introducción

Un filtro de Bloom es una estructura de datos probabilística que nos permite determinar si un elemento pertenece o no a un conjunto. Uno de los inconvenientes que tiene es que, por culpa de las colisiones en la función de hash, podríamos tener un falso positivo para un elemento. El objetivo de este trabajo ha sido hacer una análisis de los falsos positivos obtenidos, utilizando diferentes métodos de representación para el conjunto de claves; y observar si tenemos alguna diferencia significativa.

2 Implementación

2.1 Filtro de Bloom

Hemos basado nuestra implementación del filtro en un diseño que encontramos en internet, pero tuvimos que adaptarlo ya que requería también como parámetro, además del número de entradas, el error que se quería obtener. TODO: explicarlo bien y murmurhash

2.2 Generador de Strings

Para automatizar la generación de claves para los experimentos, hemos añadido también un pequeño programa que crea aleatoriamente n strings de una longitud determinada. Como, al tratarse de generación aleatoria, la muestra podría no ser representativa, también hemos utilizado un fichero de contraseñas frecuentes, y las hemos ordenado por longitud para respetar la restricción de d carácteres alfanuméricos del enunciado.

2.3 Funciones Criptográficas

Como funciones criptográficas hemos utilizado el SHA256 tal y como decía el enunciado. También hemos probado MD5, aunque pueda incluir alguna colisión de hash; pero esto es poco probable.

3 Experimentos

4 Conclusiones