Función de hasheo

Tercera entrega 25/10/2019

Ruta en GitLab: /Actividades/ADA03020/

Una función de hasheo es una función H(x) : A -> B donde B puede ser igual a A pero generalmente es un conjunto con una cardinalidad menor. Generalmente, H(x) tiene suficiente información para determinar si una preimagen coincide con una imagen pero no la suficiente para determinar x a partir de H(x). En otros términos, no debería existir H^-1.

Las funciones de hasheo tienen distintos usos, y generalmente cada función está diseñada para un uso específico. Por ejemplo BCrypt, usada en el SLTA como medio de seguridad, está diseñada para almacenar contraseñas, o más bien sus hashes, a finalidad de que el resultado se pueda utilizar para verificar una contraseña pero no para determinarla a partir del hash. Además, es criptograficamente escalable, ya que aplica un número de iteraciones de Blowfish que puede ser configurada por el programador, de modo que una vez que la velocidad de cómputo de un atacante aumente y se vuelva factible la fuerza bruta, es posible aumentar la cantidad de iteraciones y dificultar la aplicación del método.

A menudo las funciones de hasheo también son usadas para implementar estructuras de datos en las cuales la información se accede utilizando un valor no numérico como índice. Este es el caso de los diccionarios, que utilizan el hash de su objeto clave como índice en un arreglo (en implementaciones simples; otras implementaciones pueden usar árboles binarios etc). En el caso del SLTA, como parte del sistema de traducción se diseñó una función de Hash llamada KDHash, para ser utilizada para identificar los strings internacionalizables.

Existe algo llamado *ataque de colisión*, en el cual un atacante busca un valor que al aplicársele la función de hasheo devuelve el mismo valor que otro valor que es el que el sistema esperaba. Hace algunos años ocurrió con SHA1 y SHA2, demostrado por Google al almacenar en su repositorio dos archivos con distinto contenido que se computaban al mismo hash, lo cual causó errores de funcionamiento en su repositorio de Git.