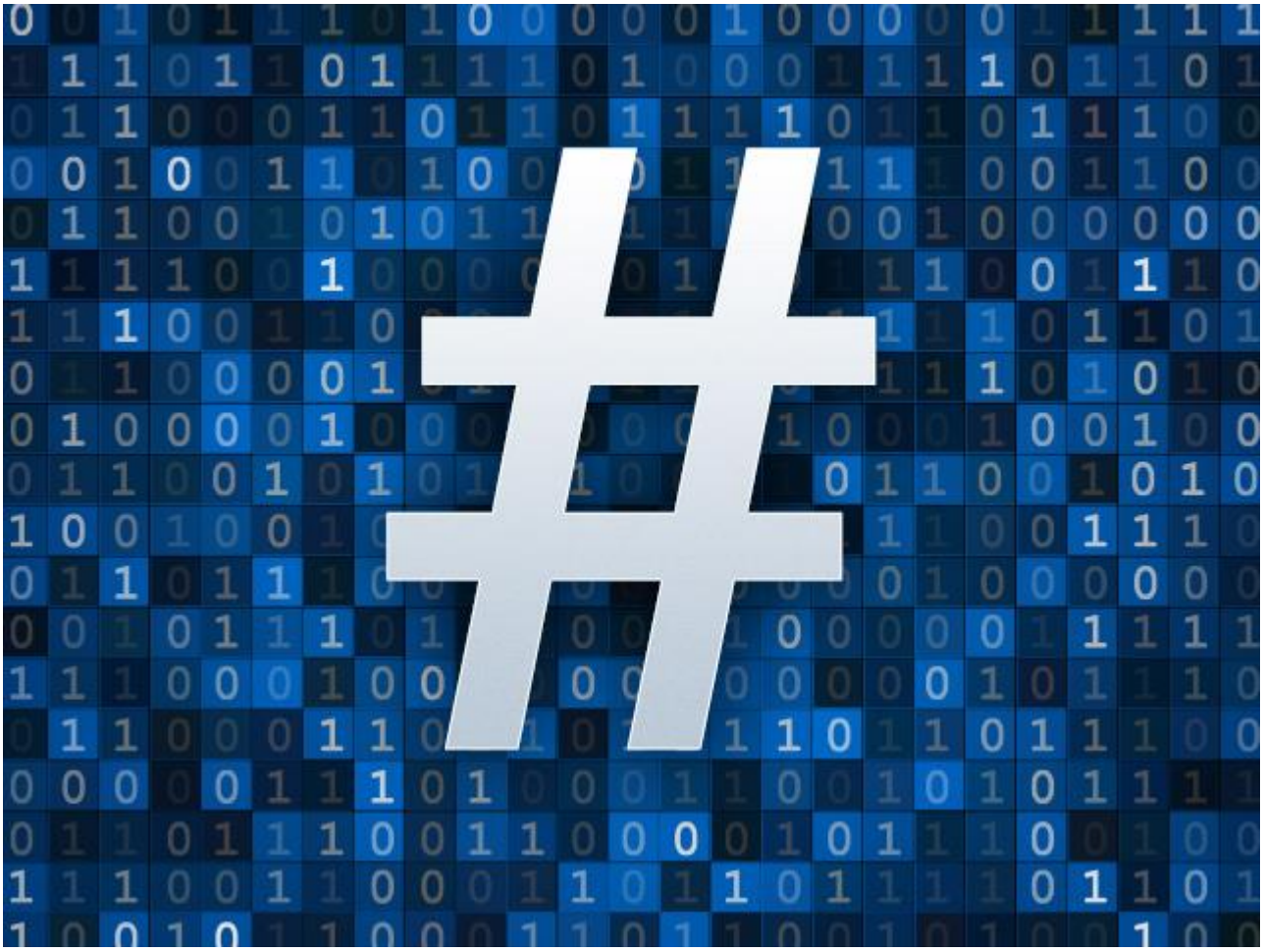
¿Qué Es Un Hash Y Cómo Funciona?

Una función criptográfica hash- usualmente conocida como “hash”- es un algoritmo matemático que transforma cualquier bloque arbitrario de datos en una nueva serie de caracteres con una longitud fija. Independientemente



Una función criptográfica hash- usualmente conocida como “hash”- es un algoritmo matemático que transforma cualquier bloque arbitrario de datos en una nueva serie de caracteres con una longitud fija. Independientemente de la longitud de los datos de entrada, el valor hash de salida tendrá siempre la misma longitud.

Según un generador online de hash SHA-1, (SHA-1 es una de las funciones hash más ampliamente utilizadas en la informática, junto con MD 5 y SHA-2) el hash de mi nombre, Brian, es: 75c450c3f963befb912ee79f0b63e563652780f0. Como probablemente cualquier otro Brian te lo puede decir, es muy común que muchas personas cometan el error de escribir “Brain” (cerebro, en inglés) cuando intentan escribir mi nombre. Es tan común, de hecho, que una vez tuve una licencia de conducir con el nombre, “Brain Donohue,” pero eso es otra historia. El hash SHA-1

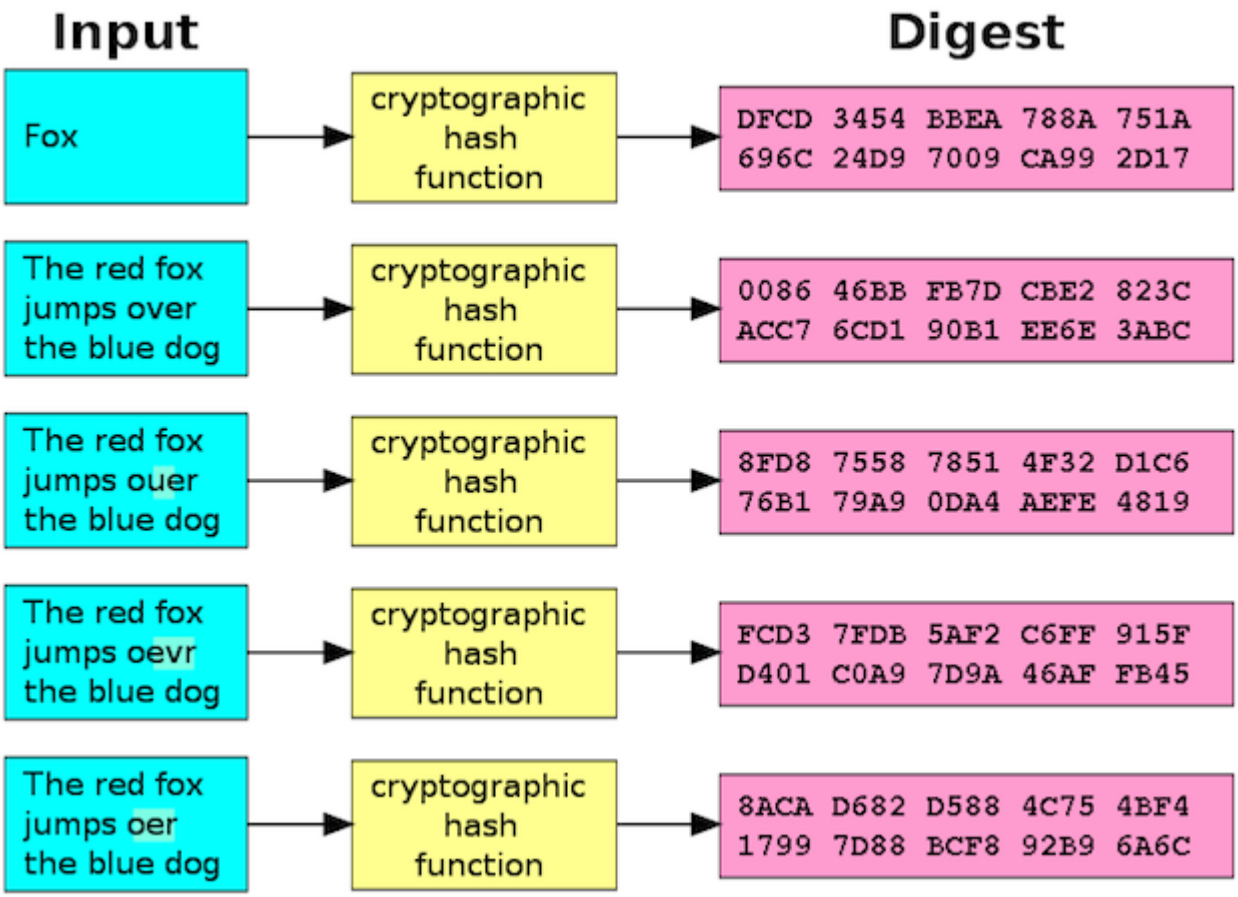
para “brain” (cerebro), siempre según el generador online de hash SHA-1, es: 8b9248a4e0b64bbccf82e7723a3734279bf9bbc4.

Como puedes ver, estas dos series son muy diferentes, a pesar de que la diferencia entre el nombre “Brian” y la palabra “brain”, radica en la disposición de dos vocales consecutivas (‘ia’ y ‘ai’, respectivamente). Yendo un poco más lejos, si introduzco mi nombre sin la “B” mayúscula, el generador de SHA-1 vuelve de nuevo con un hash completamente diferente: 760e7dab2836853c63805033e514668301fa9c47.

Te darás cuenta de que todos los hashes aquí tienen 40 caracteres de longitud, lo cual no es sorprendente dado que, en ambos casos, la entrada es de cinco caracteres. Sin embargo, lo más llamativo es que, si ingresamos todas las palabras que ya llevamos escritas en este artículo, el generador de hash arroja lo siguiente: db8471259c92193d6072c51ce61dacfdda0ac3d7.

Son 1.937 caracteres (con espacios incluidos) condensados, al igual que Brian y Brain, en una serie de 40 caracteres. Si quisieras, podrías resumir las obras completas de William Shakespeare con SHA-1 y el hash resultante seguiría siendo de 40 caracteres. Por otra parte, no existen dos entradas que produzcan el mismo hash de salida.

Aquí hay una foto, cortesía de Wikimedia Commons, que ilustra este concepto para aquellos que prefieren el aprendizaje visual:



¿Para qué se utilizan los valores hash?

Es una excelente pregunta. Sin embargo, la respuesta es que los hashes criptográficos se utilizan para un montón de cosas.

Para mí y para ti, la forma más común del hash tiene que ver con las contraseñas. Por ejemplo, si alguna vez olvidas tu contraseña de algún servicio en línea, probablemente tengas que resetearla. Cuando se restablece una contraseña, por lo general no recibes una clave en texto plano. Eso es debido a que los servicio en línea no almacenan las contraseñas en texto plano, sino que las almacenan bajo el valor hash de la contraseña. De hecho, el servicio (a menos que utilices una contraseña demasiado simple, que haga que el valor hash sea ampliamente conocido) no tiene idea de cuál es la contraseña real.

Para que quede claro, si recibes una contraseña en texto plano, quiere decir que el servicio en línea que estás utilizando no está haciendo un hashing de tu contraseña. Eso sería realmente vergonzoso.

Si lo deseas, puedes probar esto por ti mismo en un generador inverso de hash online. Si generas un valor hash de una contraseña débil (tales como “contraseña” o “123456”) y luego ingresas esa entrada en un generador inverso de hash, lo más

probable es que el generador inverso reconozca ese valor hash para cualquiera de esas dos contraseñas. En mi caso, el generador inverso de hash reconoció los hashes de ‘brain’ (cerebro) y ‘Brian’, pero no el hash que incluía todo cuerpo de este texto. Por lo tanto, se puede deducir que la integridad del hash de salida depende totalmente de los datos de entrada que, en definitiva, podría ser cualquier cosa.

Según un informe publicado en marzo en la revista TechCrunch, el popular servicio de almacenamiento de datos en la nube, Dropbox, bloqueó a uno de sus usuarios porque compartía contenido protegido por la ley estadounidense de derechos de autor (Digital Millennium Copyright Act). Este usuario luego escribió en su cuenta de Twitter que le habían bloqueado su cuenta de Dropbox por haber compartido ciertos contenidos; poco a poco se empezó a hablar de la noticia en la red social y mucha gente comenzó a quejarse de Dropbox accedía a las cuentas de los usuarios, a pesar de que en sus políticas de privacidad, declaraba lo contrario.

Dropbox, claramente, no accedió a los contenidos de este usuario. Como se lee en el artículo de TechCrunch, probablemente lo que pasó fue esto: quienes tenían los derechos de autor sobre el archivo compartido (probablemente la versión digital de una canción o de una película) crearon una función hash. Luego agregaron este hash de 40 caracteres en una especie de “lista de materiales protegidos rechazados”. Y, cuando el usuario intentó compartir el archivo, el sistema de análisis de Dropbox detectó automáticamente el hash que estaba incluido en este listado y bloqueó el intento de compartir el contenido.

Entonces, se pueden crear hash para contraseñas y archivos multimedia, pero ¿Para que otras cosas sirven las funciones criptográficas hash? Como dije antes, las funciones hash sirven para muchas más cosas de las que sé, entiendo o me interesan para los propósitos de este artículo. De todas formas, existe un uso de los hash que sí me interesa mucho: las empresas de antivirus como Kaspersky Lab, utilizan las funciones criptográficas hash para detectar malware.

Tal y como las compañías discográficas y cinematográficas crean listas de hash rechazados para proteger sus datos, también existen listas de hash de malware y muchas de ellas son de acceso público. Estas listas, o firmas de malware (como también se las conoce), están conformadas por valores hash de las piezas de malware o de algunas partes pequeñas y reconocibles de estas piezas. Por un lado, si un usuario detecta una archivo sospechoso, puede consultar en las bases de datos públicas y, de esta forma, sabrá si se trata o no de un archivo malicioso. Por el otro, los productos antivirus detectan y bloquean los malware comparando los hash con sus propias bases de datos y con las públicas.

Por otra parte, las funciones criptográficas hash se utilizan también para asegurar la “integridad de los mensajes”. En pocas palabras, para estar seguros de que algunas comunicaciones o archivos no fueron alterados de alguna forma, se

pueden examinar los hash creados antes y después de la transmisión de los datos. Si los dos hash son idénticos, significa que no ha habido ninguna alteración.