Содержание:

Введение

Криптографические хеш-функции — незаменимый и повсеместно распространенный инструмент, используемый для выполнения целого ряда задач, включая аутентификацию, проверку целостности данных, защиту файлов и даже обнаружение зловредного ПО. Существует масса алгоритмов хеширования, отличающихся криптостойкостью, сложностью, разрядностью и другими свойствами. Считается, что идея хеширования принадлежит сотруднику IBM, появилась около 50 лет назад и с тех пор не претерпела принципиальных изменений. Зато в наши дни хеширование обрело массу новых свойств и используется в очень многих областях информационных технологий.

Что такое хеш и как он работает?

Если коротко, то криптографическая хеш-функция, чаще называемая просто хешем, — это математический алгоритм, преобразовывающий произвольный массив данных в состоящую из букв и цифр строку фиксированной длины. Причем при условии использования того же типа хеша длина эта будет оставаться неизменной, вне зависимости от объема вводных данных. Криптостойкой хеш-функция может быть только в том случае, если выполняются главные требования: стойкость к восстановлению хешируемых данных и стойкость к коллизиям, то есть образованию из двух разных массивов данных двух одинаковых значений хеша. Интересно, что под данные требования формально не подпадает ни один из существующих алгоритмов, поскольку нахождение обратного хешу значения — вопрос лишь вычислительных мощностей. По факту же в случае с некоторыми особо продвинутыми алгоритмами этот процесс может занимать чудовищно много времени.

Например, мое имя — Brian — после преобразования хеш-функцией SHA-1 (одной из самых распространенных наряду с MD5 и SHA-2) при помощи онлайн-

генератора будет выглядеть так: 75c450c3f963befb912ee79f0b63e563652780f0. Как вам скажет, наверное, любой другой Брайан, данное имя нередко пишут с ошибкой, что в итоге превращает его в слово brain (мозг). Это настолько частая опечатка, что однажды я даже получил настоящие водительские права, на которых вместо моего имени красовалось Brain Donohue. Впрочем, это уже другая история. Так вот, если снова воспользоваться алгоритмом SHA-1, то слово Brain трансформируется в строку 97fb724268c2de1e6432d3816239463a6aaf8450. Как видите, результаты значительно отличаются друг от друга, даже несмотря на то, что разница между моим именем и названием органа центральной нервной системы заключается лишь в последовательности написания двух гласных. Более того, если я преобразую тем же алгоритмом собственное имя, но написанное уже со строчной буквы, то результат все равно не будет иметь ничего общего с двумя предыдущими: 760e7dab2836853c63805033e514668301fa9c47.

Заключение

Впрочем, кое-что общее у них все же есть: каждая строка имеет длину ровно 40 символов. Казалось бы, ничего удивительного, ведь все введенные мною слова также имели одинаковую длину — 5 букв. Однако если вы захешируете весь предыдущий абзац целиком, то все равно получите последовательность, состоящую ровно из 40 символов: c5e7346089419bb4ab47aaa61ef3755d122826e2. То есть 1128 символов, включая пробелы, были ужаты до строки той же длины, что и пятибуквенное слово. То же самое произойдет даже с полным собранием сочинений Уильяма Шекспира: на выходе вы получите строку из 40 букв и цифр. При всем этом не может существовать двух разных массивов данных, которые преобразовывались бы в одинаковый хеш.