НАХОЖДЕНИЕ КОЛЛИЗИИ В АЛГОРИТМЕ ХЕШИРОВАНИЯ MD5 И СПОСОБ ПОСТРОЕНИЯ МАТЕРИАЛА

А.Н.Ивкин

Самарский государственный университет

Статья посвящена наиболее интересным атак на алгоритм MD.5 и их программным реализациям, также в статье предоставляется способ построения материала, программная реализация визуально незаметных изменений документа до получения необходимого хеша.

Введение

Криптографические функции хеширования получают на вход данные произвольного размера и преобразуют их в фиксированную битовую строку. Незначительные изменения данных приводит к абсолютному изменению хеша.

Хеш-функции применяются в электронных *цифровых подписях, кодах аутентичности сообщений* и в других формах *аутентификации* в целом.

Коллизией функции хеширования называются различные блоки данных, таких, что их хеш совпадет.

Алгоритм хеширования MD5 – 128 битный алгоритм хеширования, построенный на распространенном итерационном методе Меркла-Дамгарда, на котором реализовано множество современных алгоритмов. Можно утверждать, что многие достижения в криптоанализе алгоритмов хеширования и успешные атаки на их применения были сделаны благодаря алгоритму *MD5* и его популярности, весьма привлекавшей криптоаналитиков со всего мира.

Цель данной работы, реализовать модель атаки на полноценный алгоритм хеширования MD5 в виде компьютерной программы и представить способ построения материала для упрощенного варианта известного алгоритма хеширования.

Теоретическая часть

В 2014 году была опубликована работа Ксиаоюн Ванг и ее коллег. В данной работе авторы привели два примера коллизий для двухблочных сообщений (в которых разность первых блоков двух сообщений вносит незначительное различие в состояние алгоритма, которое впоследствии аннулируется при обработке незначительно различающихся вторых блоков; схема двухблочной коллизии приведена на рис. 1) с использованием стандартного вектора инициализации алгоритма *MD5*. Детали атаки на *MD5* опубликованы не были, но приведенных примеров коллизий было достаточно для доказательства существования математических методов поиска коллизий для *MD5*.

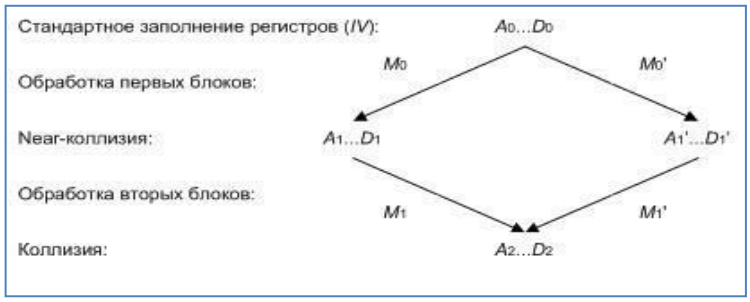

Рисунок 1 – Схема двухблочной коллизии

Несколько позже вышла работа, в которой Ксиаоюн Ванг и Хонгбо Ю описали детали атаки по поиску коллизий для *MD5*. Китайские специалисты использовали комбинированный криптоанализ – с разностью, определяемой операцией *XOR*, и с разностью, определяемой операцией вычитания по модулю 2^{32}. Как было сказано

специалистами «комбинация этих двух видов разности дает нам больше информации, чем каждый из них по отдельности».

Пример данных видов разности для двух 32-битных значений *X* и *Y*, подтверждающий их различную взаимодополняющую информативность, приведен на рис. 2.

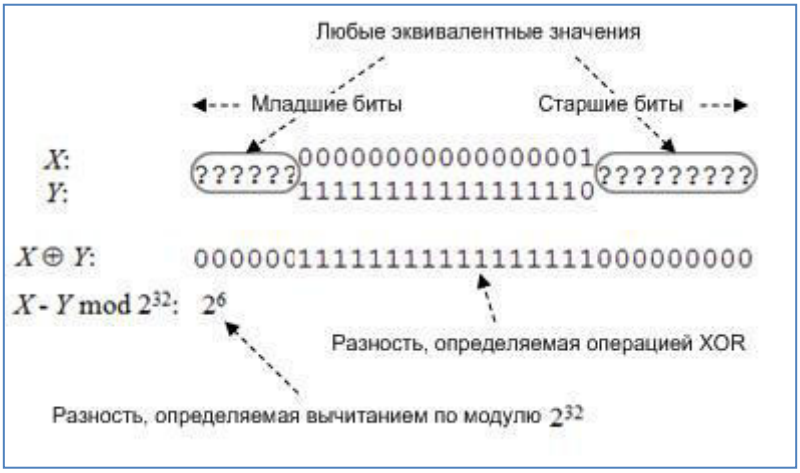

Рисунок 2 – Пример применения двух видов разности

Вначале 2006 вышла работа Властимила Климы, в которой он предложил концепцию «туннелей» в функции сжатия алгоритма *MD5*. Туннели можно представить как дополнительные условия, увеличивающие вероятность нахождения коллизии. В своей работе Клима не приводит значений трудоемкости, требуемой для нахождения коллизии с использованием туннелей, но приводит экспериментальные результаты, согласно которым среднее время нахождения коллизии для *MD5* на обычном персональном компьютере составляет 17 секунд, что однозначно являлось лучшим результатом на начало 2006 г.

В декабре 2010 г. Тао Кси и Денггуо Фенг опубликовали короткое сообщение, в котором привели пример одноблочной коллизии; формирующие коллизию сообщения имеют двухбитовую разность.

И уже в январе 2012 Марк Стивенсон опубликовал работу, в которой приведено описание атаки, а также дан пример одноблочной коллизии для алгоритма *MD5*, основывающейся на альтернативной разности сообщений:

$\Delta M = (0, 0, 0, 0, 0, 0, 0, 0, 2^{25}, 0, 0, 0, 0, 2^{31}, 0, 0)$.

Реализация

В ходе работы были реализованы атаки на алгоритм хеширования *MD*5. Для этого были написаны программы на языке программирования *C++*. Все методы реализованы в общедоступной кроссплатформенной среде разработки *CodeBlocks.*

Рассмотрим программную реализацию модели Властимила Климы и его концепции туннелей. Целью работы была работоспособная программа, которая за полиномиальное количество времени находит коллизию для алгоритма хеширования *MD5,* используя Туннели, которые можно представить как дополнительные условия, увеличивающие вероятность нахождения коллизии. В данной работе используется технология модификации сообщений предложенная Джун Яджимой и Такеши Шимоямой. На рис.3 показана работоспособность программы.

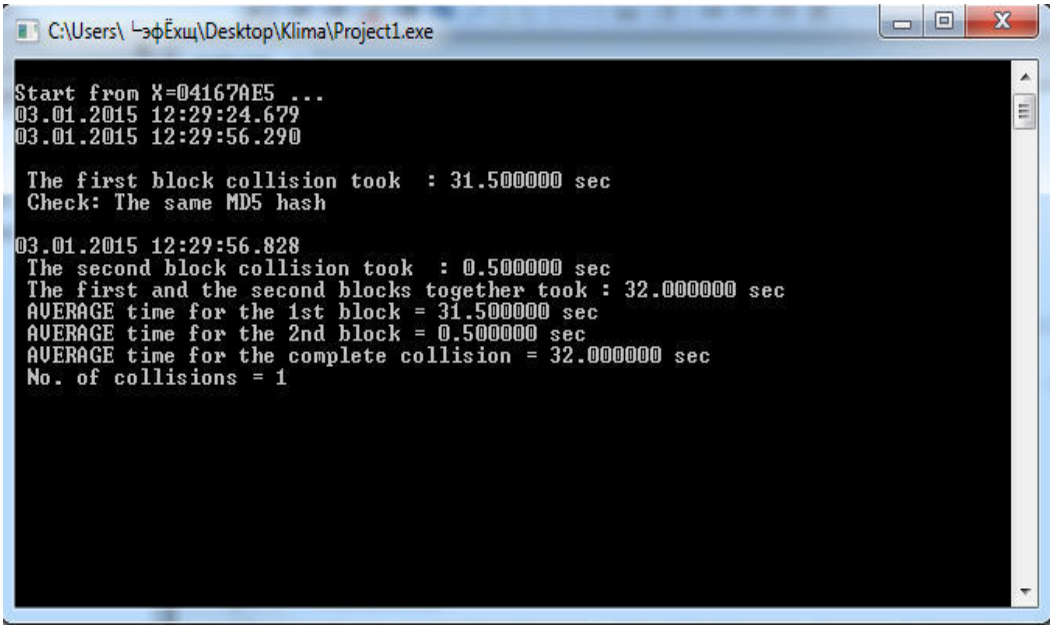

Рисунок 3 – Работоспособность программы нахождения двухблочной коллизии

Для удобства в процессе работы программы создается текстовый файл, в котором подробно описаны блоки сообщения с коллизией.

Рассмотрим программную реализацию модели Марка Стивенсона и его одноблочной коллизии. Целью работы была проверка работоспособности программы и исследование стойкости *MD5* к нахождению коллизии предложенным методом. На рис. 4 показана работоспособность программы.

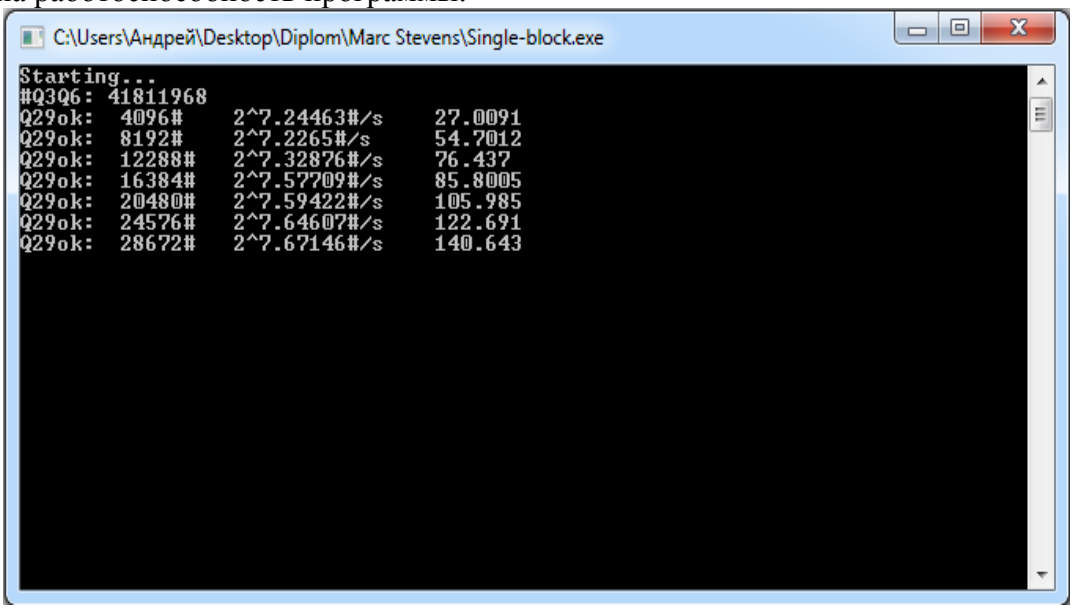

Рисунок 4 – Работоспособность программы нахождения одноблочной коллизии

Способ построения материала (экспериментальная часть)

Для приведения примера построения материала, был реализован упрощенный вариант алгоритма хеширования *MD5*. Условно назван алгоритм *L.MD5*. Данный алгоритм полностью повторяет алгоритм *MD5* за одним исключением на выходе у него первые 20 бит алгоритма. В ходе работы за основу был взят типичный договор купли продажи транспортного средства. От данного договора был подсчитан *L.MD5* хеш, после чего был создан другой текстовый файл, в котором была изменена стоимость транспортного средства. С помощью написанной программы, которая визуально незаметными изменениями (добавления пробелов в конце строк) преобразовывала хеш измененного

текстового файла до тех пор, пока он не совпадал с хешем оригинального документа. На рис.5 приведена работоспособность программы преобразования *L.MD5* хеша.

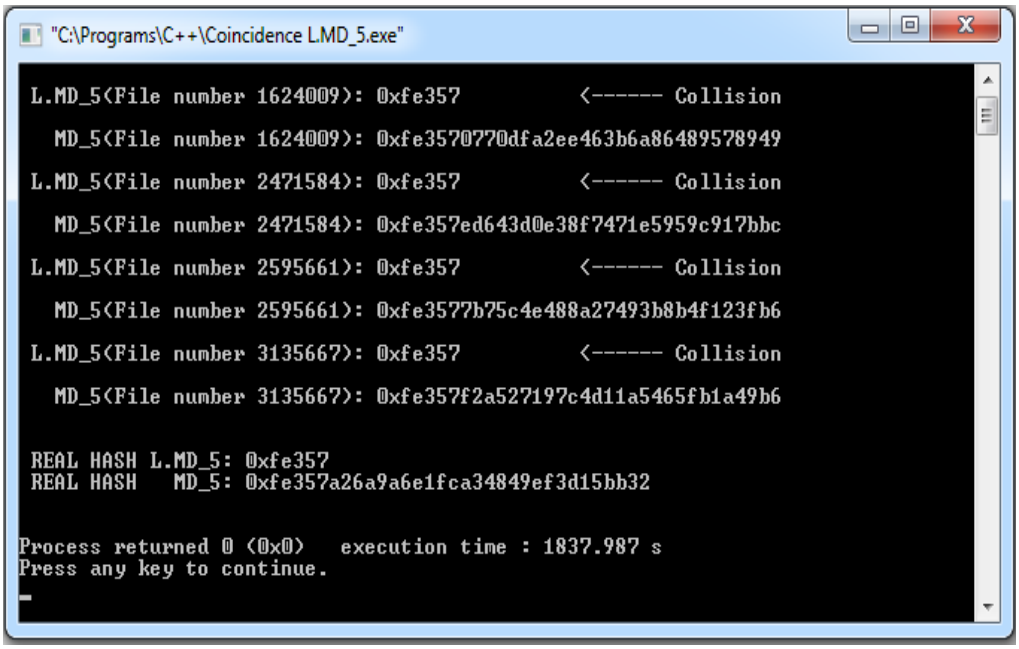

Рисунок 5 – Работоспособность программы

Заключение

При выполнении данной работы был изучен вопрос разработки хеш-функций и криптографических хеш-функций, рассмотрены существующие атаки.

Предложен способ построения материала для поиска коллизий. Программно реализована атака на упрощенный вариант алгоритма *MD5*, на языке программирования *C++*.

В ходе выполнения данной работы было изучено текущее состояние в области криптоанализа хеш-функции *MD5* в части поиска коллизий. Благодаря полученным знаниям были реализованы основные модели атак на алгоритм хеширования *MD5* в части поиска коллизий:

Все предложенные методы атак реализованы в виде компьютерных программ. Программы реализованы на языке программирования *C++*.

Литература

1. Wang X., Yu H. How to Break MD5 and Other Hash Functions. /http://citeseerx.ist.psu.edu
2. Klima V. Tunnels in Hash Functions: MD5 Collisions Within a Minute. /http://eprint.ia.cr.org
3. С.П. Панасенко публикации и научные работы /http://www.panasenko.ru/page_articles.html
4. Stevens M. Single-block collision attack on MD5. /http://marc-stevens.nl