Alexander Rhone

COMP 597 – Computer Forensics

Assignment #1

**Hypothesis:** It is hypothesized that it is feasible for a message digest algorithm to have a collision where a collision is two or more text files with different contents having the same message digest (hash value).

This hypothesis is going to be tested by creating a Java program that will change multiple characters in a row to determine if the message digest can take the original string and the modified string and output the same hash for both. The message digest (hash functions) being tested are MD5, SHA-1, and SHA-256. This program will change each of the chosen characters to each of the hexadecimal characters and later all the alphanumeric characters. There will be three input files of 10, 46, 114 characters in length. The input files will consist of numbers, letters, and punctuation. If the program takes longer than 15 minutes to finish, then it will be cancelled for fear of overheating and crashing the laptop that is being used to run the program. The laptop that is running the experiments has an Intel ® Core ™ i7-5500U CPU @ 2.40 GHz and 2.4 GHz processor, 8.00 GB of RAM, and a 64-bit Windows 10 operating system.

**First Experiment:** During the first experiment, the smallest test file was run through all 3 hash functions with 4-7 character modifications being run for each test case. This experiment is only using the hexadecimal characters. The items being tested are the number of copies created, the number of collisions that have occurred, and the time that the program has been running for.

The first text string used is:

*helloworld*

|  |  |  |  |
| --- | --- | --- | --- |
| 4 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 429807 | 429807 | 429807 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 0.729753586seconds | 0.509396837seconds | 0.793110925seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 5 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 5893871 | 5893871 | 5893871 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 10.148177339seconds | 7.640183207seconds | 9.733283013seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 6 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 78573295 | 78573295 | 78573295 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 160.524735496seconds | 119.175517675seconds | 158.686995475seconds |

After attempting the 7 modifications for 15+ minutes, all 3 of the hash functions were terminated.

There was not much to be learned from the first experiment, as expected, since no collisions were found. This is to be expected as smaller text files are more unlikely to yield collisions.

**Second Experiment:** During the second experiment, the medium test file was run through all 3 hash functions with 4-6 character modifications being run for each test case. This experiment is only using the hexadecimal characters. The items being tests are the number of copies created, the number of collisions that have occurred, and the time that the program has been running for.

The second text string used is:

*this is a text document*

*this is its second line*

|  |  |  |  |
| --- | --- | --- | --- |
| 3 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 168943 | 168943 | 168943 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 0.275624396seconds | 0.22085601seconds | 0.27237508seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 4 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 2641647 | 2641647 | 2641647 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 4.782277171seconds | 3.986033255seconds | 5.355985871seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 5 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 41283311 | 41283311 | 41283311 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 88.794261741seconds | 65.93323982seconds | 85.356861705seconds |

After attempting the 6 modifications for 15+ minutes, all 3 of the hash functions were terminated.

There was not much to be learned from this experiment as no collisions were to be found. This is also to be expected as there are only 16 modifications but was somewhat disappointing as I had thought maybe one of the larger character modifications might yield a result.

**Third Experiment:** During the third experiment, the largest test file was run through all 3 hash functions with 3-5 character modifications being run for each test case. This experiment is only using the hexadecimal characters. The items being tests are the number of copies created, the number of collisions that have occurred, and the time that the program has been running for.

The third text string used was:

*Hello this is a text document*

*Th7s i$ tQe 2nd linE*

*Th1s iS thP third l1ne*

*ThIs i8 thE fourth lINe*

*Th!s is th3 5th line*

|  |  |  |  |
| --- | --- | --- | --- |
| 3 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 430063 | 430063 | 430063 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 1.3457055seconds | 0.998693517seconds | 1.279358434seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 4 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 41283311 | 41283311 | 41283311 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 21.014497464seconds | 12.907925708seconds | 17.538411865seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 5 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 108130031 | 108130031 | 108130031 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 289.560298383seconds | 216.67861618seconds | 289.678968139seconds |

As the medium test file took longer than 15 min for 6 modifications, 6 modifications was not attempted for the largest test file.

There was not much to be learned from this experiment as no collisions were to be found. This is also to be expected as there are only 16 modifications but was disappointing as one of the larger character modifications with the largest test case was expected to cause a collision.

**Fourth Experiment:** During the fourth experiment, the smallest test file was run through all 3 hash functions with 2-4 character modifications being run for each test case. This experiment is using the alphanumeric characters. The items being tests are the number of copies created, the number of collisions that have occurred, and the time that the program has been running for.

The first text string used is:

*helloworld*

|  |  |  |  |
| --- | --- | --- | --- |
| 2 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 34037 | 34037 | 34037 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 0.070121539seconds | 0.054801269seconds | 0.066911998seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 5893871 | 5893871 | 5893871 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 2.668496455seconds | 1.994167229seconds | 2.689000744seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 4 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 78573295 | 78573295 | 78573295 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 153.920729235seconds | 116.270251546seconds | 150.235074882seconds |

There was not much to be learned from this experiment like previous experiments. This is to be expected as like the first experiment, smaller text documents are less likely to create a collision.

**Fifth Experiment:** During the fifth experiment, the smallest test file was run through all 3 hash functions with 4-7 character modifications being run for each test case. This experiment is using the alphanumeric characters. The items being tests are the number of copies created, the number of collisions that have occurred, and the time that the program has been running for.

The second text string used is:

*this is a text document*

*this is its second line*

|  |  |  |  |
| --- | --- | --- | --- |
| 2 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 170189 | 170189 | 170189 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 0.270260526seconds | 0.210617788seconds | 0.273736069seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 10317233 | 10317233 | 10317233 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 16.408419648seconds | 12.47813962seconds | 16.067049702seconds |

This experiment was disappointing because by this time at least one collision was expected to have occurred.

**Sixth Experiment:** During the sixth experiment, the smallest test file was run through all 3 hash functions with 2-3 character modifications being run for each test case. This experiment is using the alphanumeric characters. The items being tests are the number of copies created, the number of collisions that have occurred, and the time that the program has been running for.

The third text string used was:

*Hello this is a text document*

*Th7s i$ tQe 2nd linE*

*Th1s iS thP third l1ne*

*ThIs i8 thE fourth lINe*

*Th!s is th3 5th line*

|  |  |  |  |
| --- | --- | --- | --- |
| 2 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 427365 | 427365 | 427365 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 1.063927401seconds | 0.787487705seconds | 1.059830906seconds |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 modifications | Md5 | Sha1 | Sha-256 |
| Number of copies tested: | 41283311 | 41283311 | 41283311 |
| Number of collisions: | 0 | 0 | 0 |
| Elapsed time: | 66.761831043seconds | 48.702729287seconds | 70.32527263seconds |

**Final Conclusion**

This overall experiment was very disappointing because I had assumed that at least one collision was able to be accomplished. Since MD5 is considered a broken hash algorithm, I had assumed that it was easy to find collisions but after I had run these test cases, I have come to the conclusion that it is very difficult to run a brute force attack on MD5 and get at least one collision in any reasonable amount of time. I did not think that I would be able to find collisions with SHA-1 or SHA-256 as SHA-1 has only one published collision and SHA-256 has not had any published collisions.

The hypothesis in which it is feasible for a message digest algorithm to have a collision where a collision is two or more text files with different contents having the same message digest (hash value) has been refuted due to the data that has been collected.