# Project 2: Hash Attack

*Paula Chen*

## Background

This project is designed to conduct a collision attack and a second pre-image attack against a toy SHA-1 hash algorithm, to observe the relationship between different digest sizes and the number of attempts to perform a single attack, and to compare it with the theoretical numbers.

## Methodology

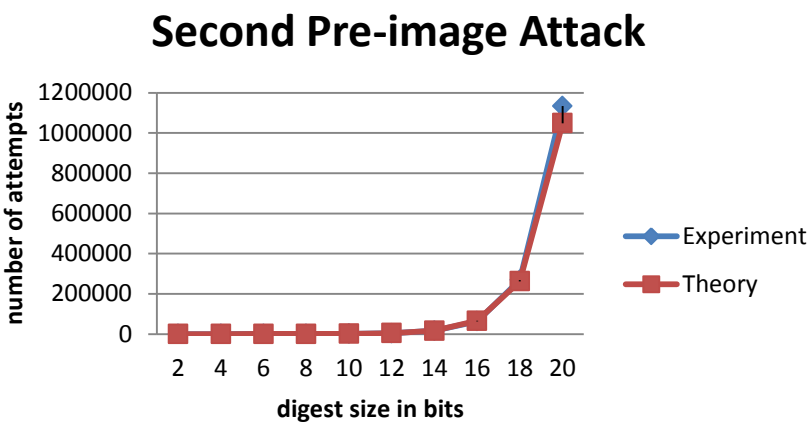
The experiments are implemented in Java using the standard Java library for the SHA-1 hash algorithm. It is modified a little bit to make the digest size adjustable by truncating the original hash into the desired number of bits.

In order to perform a second pre-image attack, a target string is first randomly generated and hashed. Then, there is a while loop that keeps generating a new random string, hashing it, and comparing it to the target hash. The only condition to break out of the while loop is when the new hash is equal to the target hash and the new string is not equal to the target string. In the meantime, the total number of attempts is being recorded until a second pre-image attack is performed.

In order to perform to a collision attack, a hashed map is first created to store randomly generated strings and their hashes. Then, there is a while loop that keeps generating a new random string, hashing it, and checking if the hash already exists in the hashed map. The only condition to break out of the while loop is when the new hash has already existed in the hashed map and the new string is not equal to the string for the existing hash. In the meantime, the total number of attempts is recorded until a collision attack is performed.

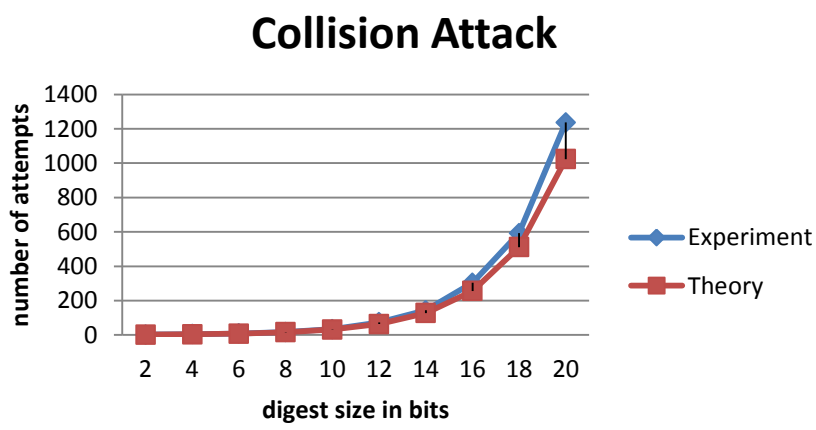
Each attack is performed 100 times to get the average number of attempts, in order to get a more accurate result. For each kind of attack, different digest size, 2, 4, 6, 8, 10, 12, 14, 16, 18, and 20 bits are experimented.

## Result




| Attempts | Experiment | Theory |
|---|---|---|
| 2 | 3.84 | 4 |
| 4 | 15.69 | 16 |
| 6 | 63.07 | 64 |
| 8 | 255.56 | 256 |
| 10 | 1013.4 | 1024 |
| 12 | 3999.56 | 4096 |
| 14 | 16554.03 | 16384 |
| 16 | 63228.6 | 65536 |
| 18 | 269002.23 | 262144 |
| 20 | 1133825.35 | 1048576 |

The graph above shows the result for the second pre-image attack, and the table shows the exact numbers. In theory, the number of attempts needed to perform an attack is $2^{(\text{number of digest size in bits})}$. We can see that the experiment line and theoretical line almost perfectly match up except for when the digest size is 20 bits. The reason for this might be that the experiment is only run for 100 times and the average is retrieved. If it is run for more times, the result is going to be more accurate.




| Attempts | Experiment | Theory |
|---|---|---|
| 2 | 2.22 | 2 |
| 4 | 4.74 | 4 |
| 6 | 9.09 | 8 |
| 8 | 17.11 | 16 |
| 10 | 33.76 | 32 |
| 12 | 74.23 | 64 |
| 14 | 143.36 | 128 |
| 16 | 304.09 | 256 |
| 18 | 592.59 | 512 |
| 20 | 1236.47 | 1024 |

The graph above shows the result for the collision attack, and the table shows the exact numbers. In theory, the number of attempts needed to perform an attack is $2^{(\text{number of digest size in bits}/2)}$. We can see that the experiment line and theoretical line don't match well after when the digest size is 12 bits. I am guessing it is not run enough times as well since the experiment line is still exponential and close to the theory line.