Seth Johnson
Dr. Scott Ruoti
9/24/2022
Reviewer: DWhite75

Testing Theoretical Costs of both Collision and Preimage Attacks on the SHA-1 Algorithm at Varying Bit Lengths

The purpose of this lab was to test the difficulty and accuracy of the theoretical costs associated with breaking the SHA-1 algorithm. A collision in hash algorithms occurs when two different plaintext pieces of data hash to the same value. A preimage attack is using one particular hash, and comparing the hashes of many random plaintext values until finding a hash that matches the hash of the original. The theoretical costs of each attack are $2^{n/2}$ for collision attacks and 2^{n} for preimage attacks where n is the number of bits in the hex digest. For example, this means for a hash produced by the SHA-1 algorithm of 8 bits long will take $2^{8/2} = 16$ attempts to find a collision and $2^{8} = 256$ attempts to successfully execute a preimage attack.

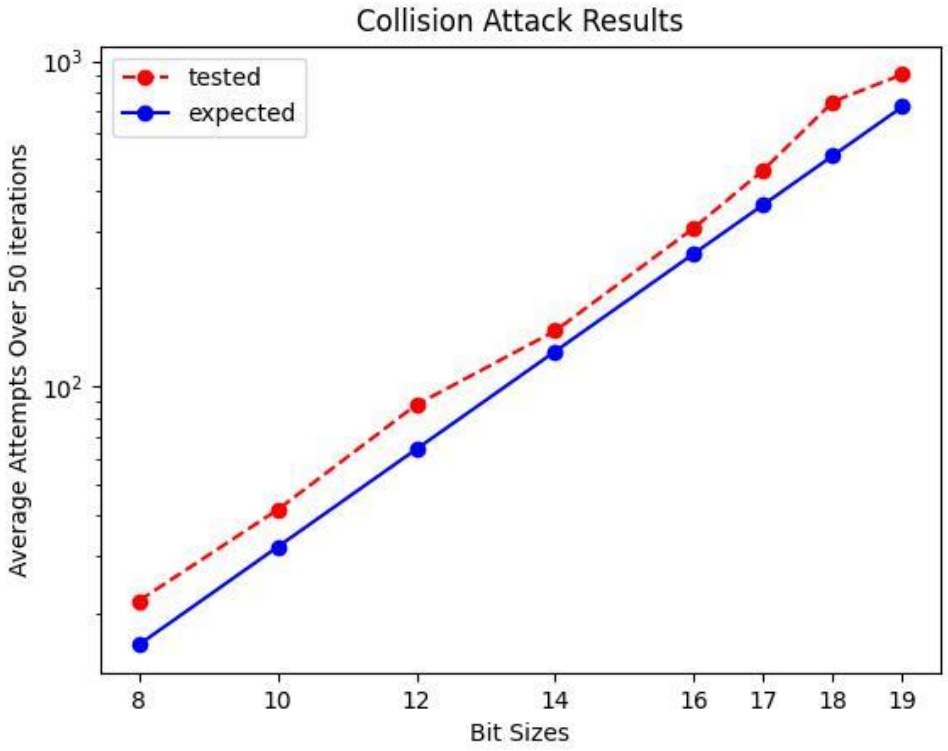
To test the accuracy of these costs, I used the SHA-1 algorithm to create hashes and truncated them to certain bit lengths. For example, a plaintext that results in the hash, "c9c4f59c8285b0308c53031acdfd158586ae59f5" would truncate to just 8 bits and result in "c9" etc. Throughout testing I used bit lengths 8, 10, 12, 14, 16, 17, 18, and 19 for both attacks to gather data. I gathered 50 samples for each bit length where there was a collision or proof of preimage occurring. For each sample I tested collision and preimage attacks by creating unique random plaintext strings, hashing them, and truncating the hash result to my desired bit length. After that, for collision attacks I stored each hash value in a set to make sure it is unique and tracked when I ran into a duplicate hash. For the preimage attack, I created one single random hash and stored it, then repeatedly hashed new random values until I found one that matched the original I stored. I wrote all the data I found to a file to record my attempts for each round and

Seth Johnson
Dr. Scott Ruoti
9/24/2022
Reviewer: DWhite75

particular bit length. I then used python to read in the data and create graphs using matplotlib.

The results are shown below.

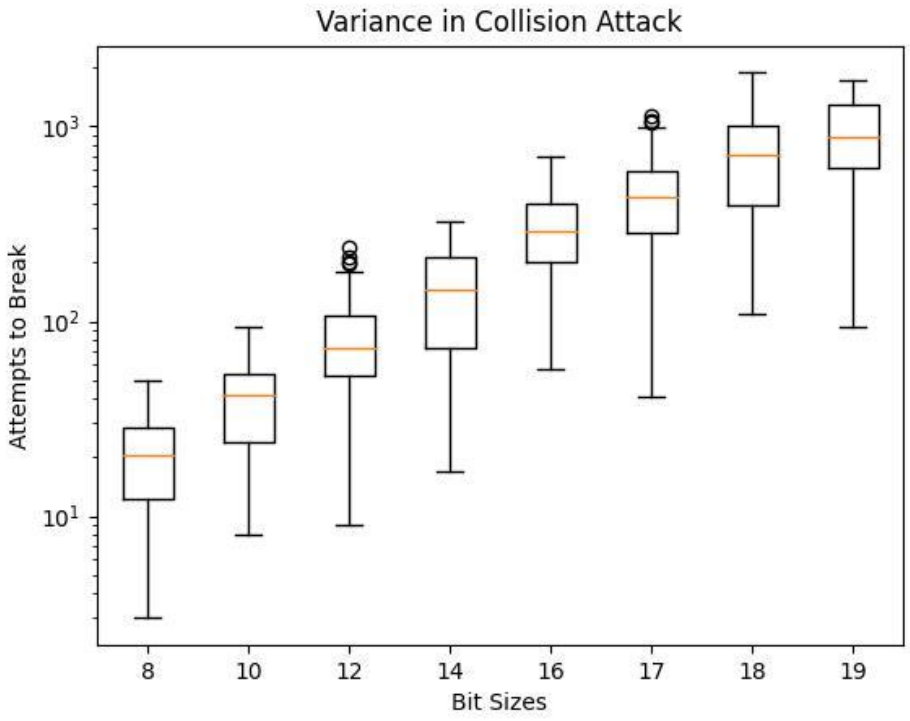
Collision Attack:



Bit Size	8	10	12	14	16	17	18	19
Theoretical Cost	16	32	64	128	256	362	512	724
My Avg. Attempts	21	41	87	148	307	463	749	910

Seth Johnson
Dr. Scott Ruoti
9/24/2022
Reviewer: DWhite75

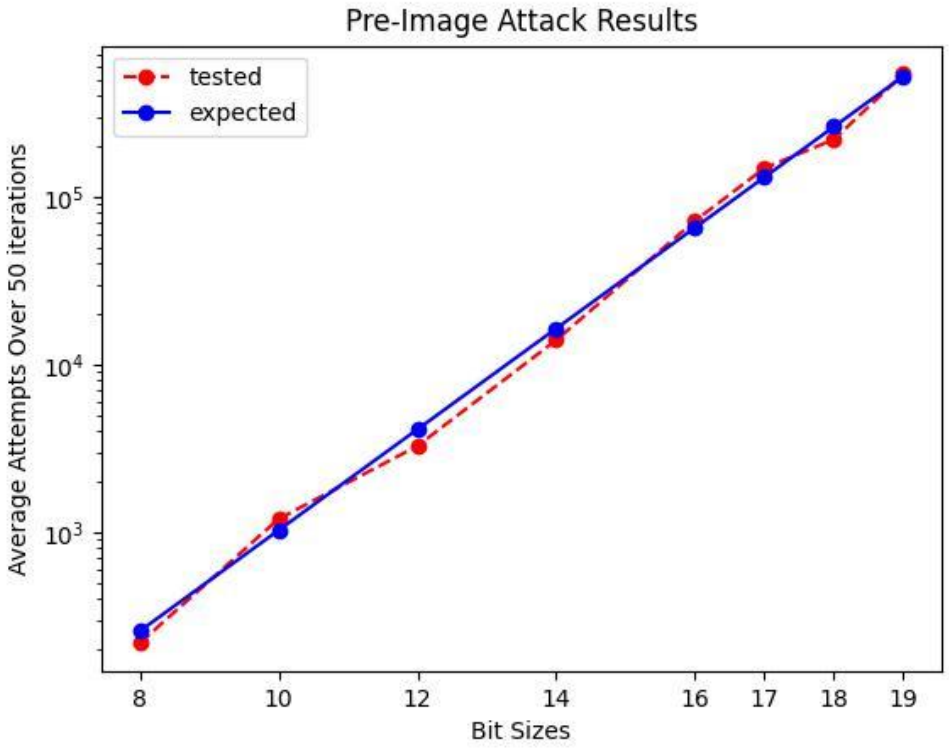
As can be seen from the data, my data is fairly consistent with the theoretical costs. My expected reason for the difference is that there are outliers within the data that can change the average and make quite an impact on the data. This can be seen by the variance box and whisker graph below.



The yellow line represents the median values of all the attempts for each bit length and the dots represent outliers that are outside the high and low frequencies. This chart better shows the range of attempt values throughout the bit values. However, even with those outliers the results were fairly close to the theoretical.

Seth Johnson
Dr. Scott Ruoti
9/24/2022
Reviewer: DWhite75

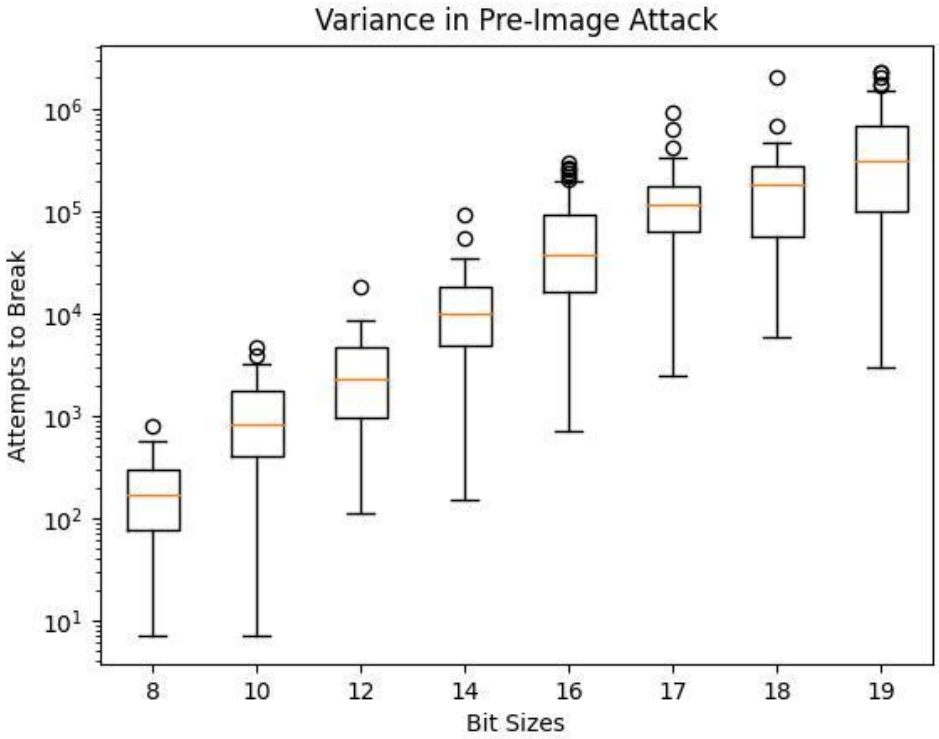
Preimage Attack:



Bit Size	8	10	12	14	16	17	18	19
Theoretical Cost	256	1,024	4,096	16,384	65,536	131,072	262,144	524,288
My Avg. Attempts	217	1,196	3,261	13,978	71,959	148,452	220,042	540,244

As can be seen by the graphs, the results of the preimage graphs are closer to the theoretical than the collision results. I suspect this to be because preimage naturally requires much more attempts than collision. The variance in the results can be seen below from the box and whisker graph.

Seth Johnson
Dr. Scott Ruoti
9/24/2022
Reviewer: DWhite75



According to the graph above, there are more outliers within this attack compared to the collision results.

Both the outcomes of the attacks were as expected based on the theoretical costs. In total, to test each bit length 50 times, it took the collision attack around 30seconds to a minute to test, and 18-20minutes to test. With SHA-1 normally producing 160-bit length hashes, it would take 1.4615016e+48 attempts just to prove one preimage and 1.2089258e+24 attempts to show a collision. This has been proven to be an insecure algorithm as SHA-1 has been broken and other algorithms in the SHA family are proven to be safer for use.