Joshua Van Steeter
CS 465
Project 2 Write Up

Hash Attack

Purpose:

To determine the number of hashes needed to be randomly generated in order to find another string that hashed to the same value, at varying different bit length of hashes.

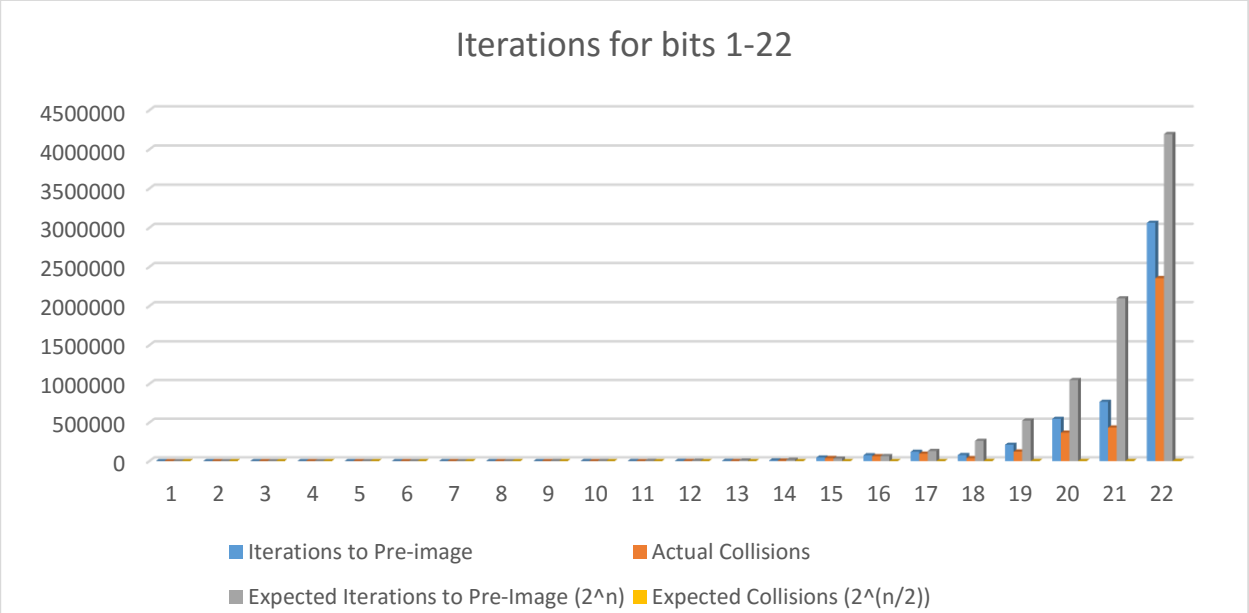
Procedure:

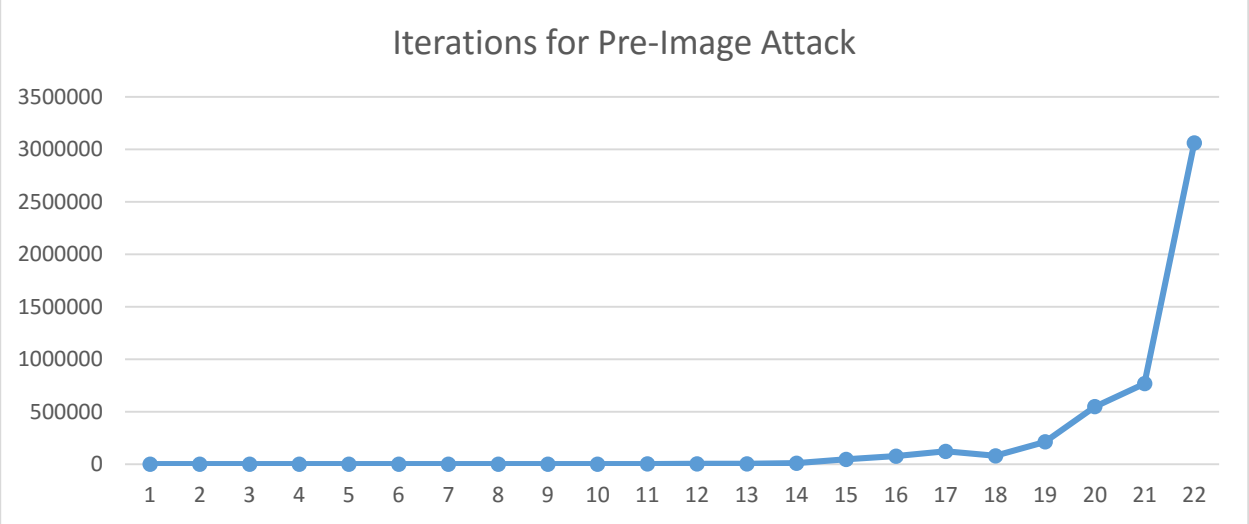
Using the Java class, SecureRandom, I randomly generated strings which I then converted into a SHA-1 hash. I then stored the first n bits of the hash into a HashSet. The first randomly generated hash was the "pre-image", meaning that the test would continue hashing random strings until another hash was produced that had the same hash as the pre-image. While hashes where being produced in an attempt to find a hash that matched the pre-image hash, any time a hash was produced that matched any existing hash in the set a "collision" was recorded. When a hash matching the pre-image hash was discovered, the number of iterations it took to find a duplicate hash was recorded along with the number of collisions that were recorded in the meantime.

Data:

bits	Iterations to Pre-image	Actual Collisions	Expected Iterations to Pre-Image (2^n)	Expected Collisions (2^(n/2))
1	1	1	2	1.4142136
2	1.76	1.3	4	2
3	3.98	2.38	8	2.8284271
4	7.54	3.9	16	4
5	50.94	42.14	32	5.6568542
6	30.76	14.98	64	8
7	71	40.98	128	11.313708
8	150.46	98.22	256	16
9	188.44	88.5	512	22.627417
10	469.46	263.2	1024	32
11	652.84	286.3	2048	45.254834
12	2078.56	1194.8	4096	64
13	3378.74	1822.62	8192	90.509668
14	8595.46	5265.06	16384	128
15	46108.4	39387.16	32768	181.01934
16	75700	63291.3	65536	256
17	121065	96327.84	131072	362.03867
18	78550.14	38944.6	262144	512
19	211676.8	123429.22	524288	724.07734
20	546850.88	368215.5	1048576	1024
21	766669.02	435188.2	2097152	1448.1547

22	3061911.68	2355699.02	4194304	2048





Conclusion:

Between 1 and 10 bits the time to find a matching pre-image is negligible. After about 14 bits the number of iterations it took for a successful pre-image attack began to noticeable increase at an exponential rate. The actual number of collisions found before a successful pre-image attack was significantly higher than the expected collisions. The actual pre-image iterations were back and forth with the expected iterations, but overall still significantly lower. Suggesting either it is much easier to find collisions than predicted, or the hash generating algorithm isn't as diverse as expected.