Additional Evaluation Details

This document provides additional details and results for the evaluation of the attacks on the KMV sketch presented in the paper: *P. Reviriego, A. Sánchez-Macian, S. Liu and F. Lombardi "On the Security of the K Minimum Values (KMV) Sketch", under submission to IEEE Transactions on Dependable and Secure Computing.*

1) Simulation details and data sets

For the experiments, randomly generated elements of type double from a uniform distribution have been used. For the inflation attack, sets of different cardinalities (10000, 28480, 81113, 231013, 657933, 1873817, 5336699, 15199111, 43287613, 123284674, 351119173, 1000000000), C_{KMV} in the paper, have been built using these random numbers. Subsets of them have been extracted to create the reduced attack set of cardinality C_A. For the deflation attack, sets of 10 million of uniformly distributed randomly generated double values have been originally generated to obtain an attack set of size 100,000. Parameter "t", defined in the paper, has been tested for different values (100, 1514, 2929, 4343, 5757, 7171, 8586, 10000). However, if the hash functions are well behaved, then any practical data set will produce similar results; hash functions are tested to ensure that they generate results equivalent to a random mapping for different data sets (see for example, https://github.com/rurban/smhasher). To show that this is the case, some of the experiments have been run with a sequential data set made of increasing integer numbers. The results are nearly the same as when using random input elements. As an example, the results for the inflation attack are shown below in Figure 1 for random data and sequential data, they are almost identical. The code in the repository is able to select between random or sequential elements.

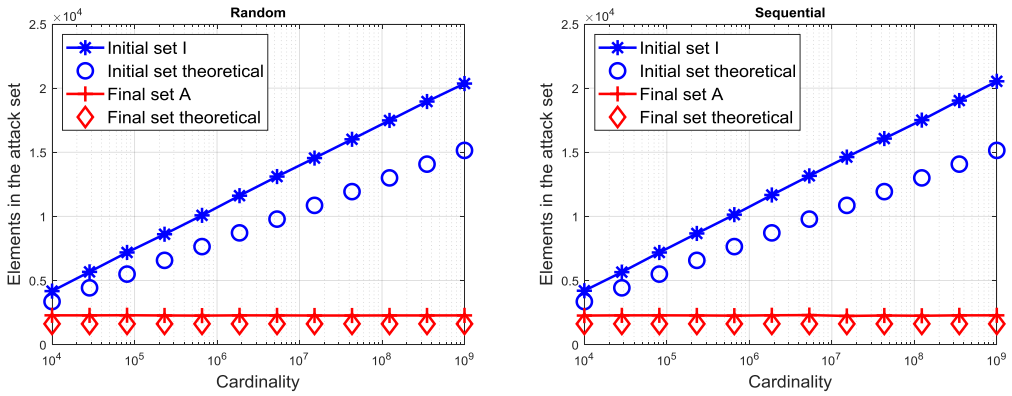

Figure 1 Inflation attack on the Apache DataSketches library KMV implementation K = 1024 using random data (left) and sequential data (right)

2) Evaluation of different values of K

The paper presents results for a single value of K. As K is a critical parameter that determines the size and accuracy of the sketch it is of interest to check if it has any impact on the attack effectiveness. However, the effectiveness of the proposed attacks is not dependent on the value of K. To show that this is the case, some of the experiments have been run with a smaller value

K = 256 and a larger value K = 4096 for both the “ideal” implementation and the DataSketches library (the code has been changed to select a value of K using command line arguments). The results are shown in the Figures below; they are like those for K = 1024. The only thing worth mentioning is that the values seem to increase over the upper bound because K increases for the deflation attack.

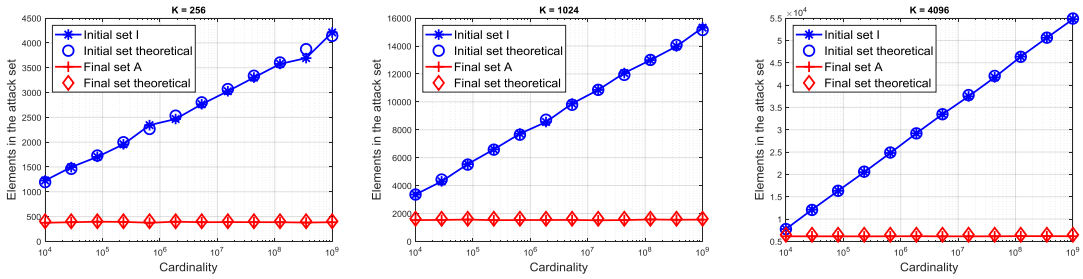

*Figure 2 Inflation attack on the ideal KMV implementation for K = 256 (left), K = 1024 (middle) and K = 4096 (right). It can be seen that the attack is efficient in all cases and the attack set size is O(K) for the final set A and O(K)*log(C) for the initial attack set I*

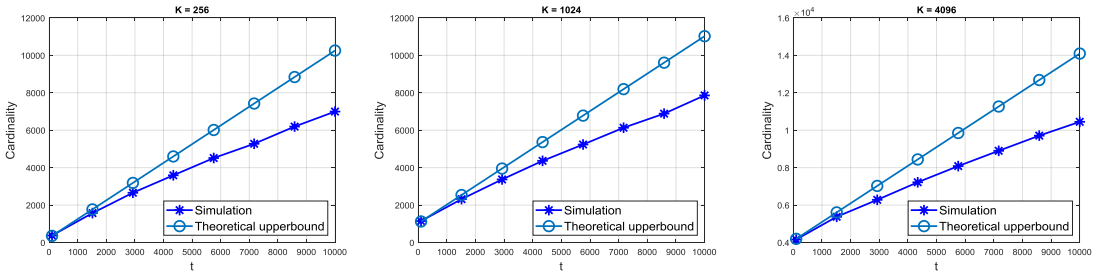

Figure 3 Deflation attack on the ideal KMV implementation for K = 256 (left), K = 1024 (middle) and K = 4096 (right). Results are similar to those of K = 1024 with the attack set KMV cardinality estimate being much smaller than the actual set cardinality

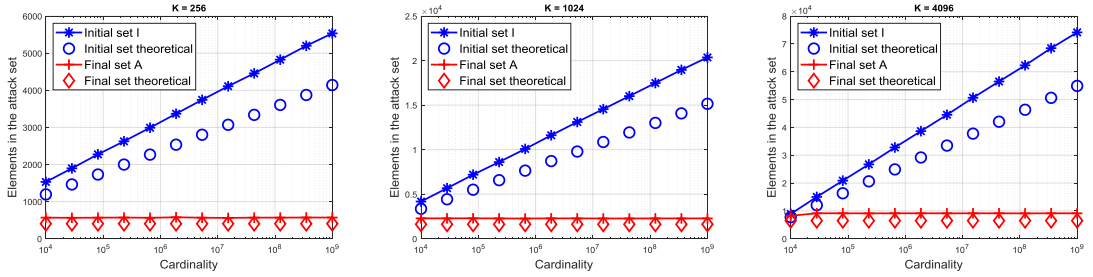

Figure 4 Inflation attack on Apache DataSketches KMV implementation for K = 256 (left), K = 1024 (middle) and K = 4096 (right)

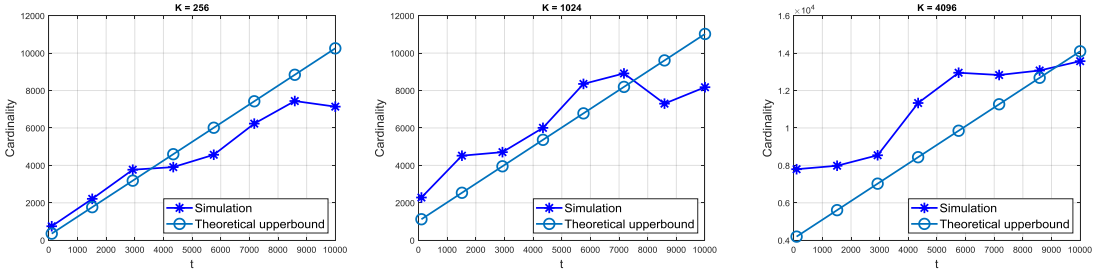

Figure 5 Deflation attack on the Apache DataSketches l KMV implementation for K = 256 (left), K = 1024 (middle) and K = 4096 (right)