Internet Data Streaming Project 3 - Implementation of Flow Size Sketches

CountMin : In order to calculate the flow size, we are hashing the flow ids and incrementing the counter array positions and later we are trying to get the minimum of all the respective counter indexes for the flow ids. We are doing this to reduce the error rate. Parse Class is used to take the input from a input file which has the number of flow ids followed by flow ids, k counters, counter size w. CountMin Class has functions callback() which is responsible for calculating hashes for the read flow ids into k counters of w size each and findMinimum class is responsible to calculate the minimum of all hashed counter index values of flow ids to calculate the flow size in order to reduce the error rate.

Demo Input : 10000 3 300

Output : error rates : 127
original, flowids, expected
9958 1000469222 9956
9903 -2080126952 9747
9890 -25830423 8968
9855 1066611402 9835
9848 779737746 9848
9712 1427163130 9681
9665 921692844 8673
9518 1750167563 9510
9453 1806352294 9450
9339 -398712802 9324
9031 -754487550 8982
8914 -186803072 8910
8808 -815151929 8698
8761 856234373 8733
8640 713078729 8584
8492 -1379012189 8462
8422 1983357079 8390

Trimmed output

CounterSketch : In order to reduce the error rate further down, after hashing the flow ids, we only increment or update the minimum of hashed counter index values. This is more accurate than CountMin. Parse Class is used to take the input from a input file which has the number of flow ids followed by flow ids, k counters, counter size w. CounterSketch Class has functions callback() which is responsible for calculating hashes for the read flow ids into k counters of w size each and findAverageclass is responsible to calculate the median of the counter values to

record the flow size and later at the end we are recording error rates and this is more accurate than CountMin.

Demo Input : 10000 3 300
Output : error rates: 239
original, flowids, expected
10284 779737746 9848
9892 1427163130 9681
9790 -2080126952 9747
9704 1000469222 9956
9596 1066611402 9835
9508 1750167563 9510
9450 1806352294 9450
9405 -398712802 9324
9146 -25830423 8968
9014 -754487550 8982
8904 -186803072 8910
8747 856234373 8733
8698 -815151929 8698
.
.
.
Trimmed output.

Active Counters : We are trying to minimize the number of counters for long ranged numbers. Parse Class is used to take the input from a input file which has the number of flow ids followed by flow ids, k counters, counter size w. ActiveCounter class has functions callback() to calculate the hash values and it calls enter function and enter() is used to update the counter i.e to calculate the probability and later to increment the number counters accordingly i.e get the decimal value for total number of counters outside.

For the demo input given, this is the resulting output :
1005695