

INTERNET DATA STREAMING PROJECT IMPLEMENTATION OF MULTI FLOW SPREAD SKETCHES

1. VIRTUAL BITMAP

Here, we are trying to calculate the flow spread for large estimation ranges. In this I have created a class virtual bitmap and constructor to initialize the flow ids, elements, virtual bitmap size. For each flow we are maintaining a virtual bitmap and calculating the hash function for each element for flow, now we maintain a bitmap of size 500000 and virtual bitmap of size 500 and finally we are hashing for overall in the main bitmap.

The output file is called graphplot.txt

For the demo input this the output being generated :

9 20

205 207

20 72

5 44

192 198

15 22

175 198

9 46

42 44

358 379

51 76

12 20

200 188

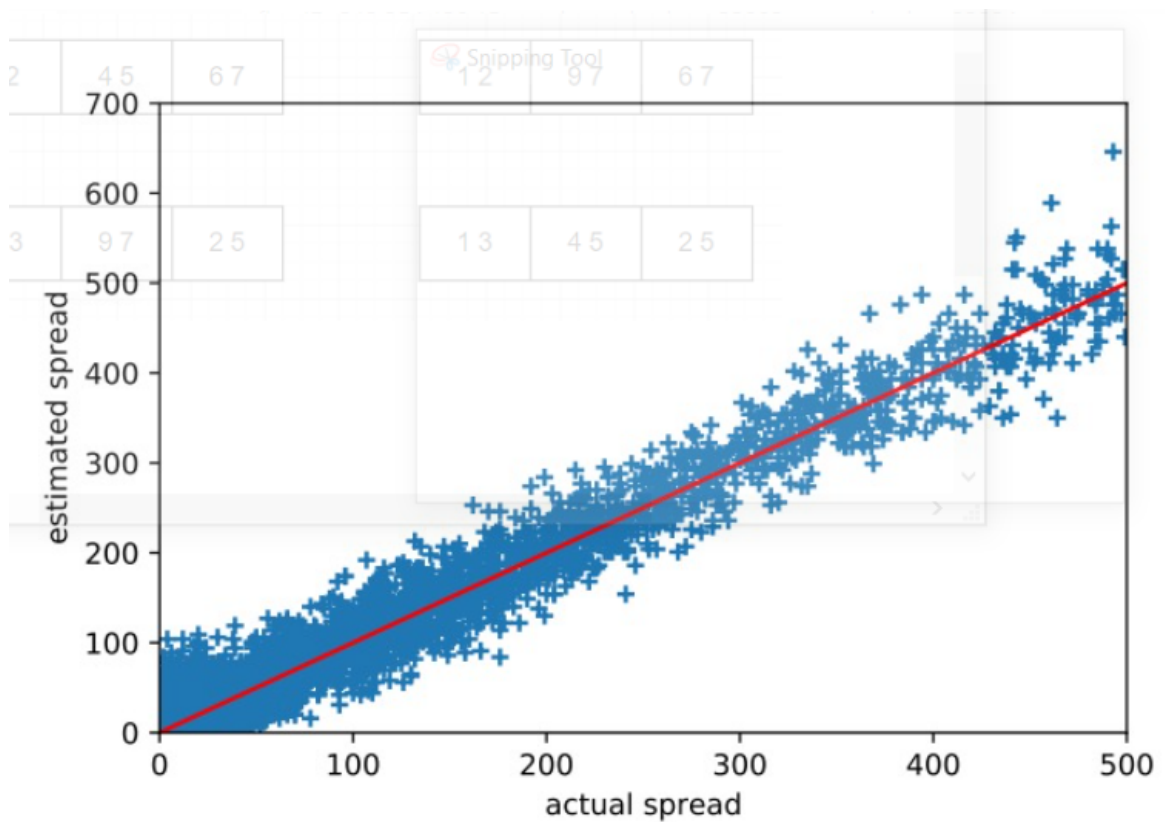
103 96

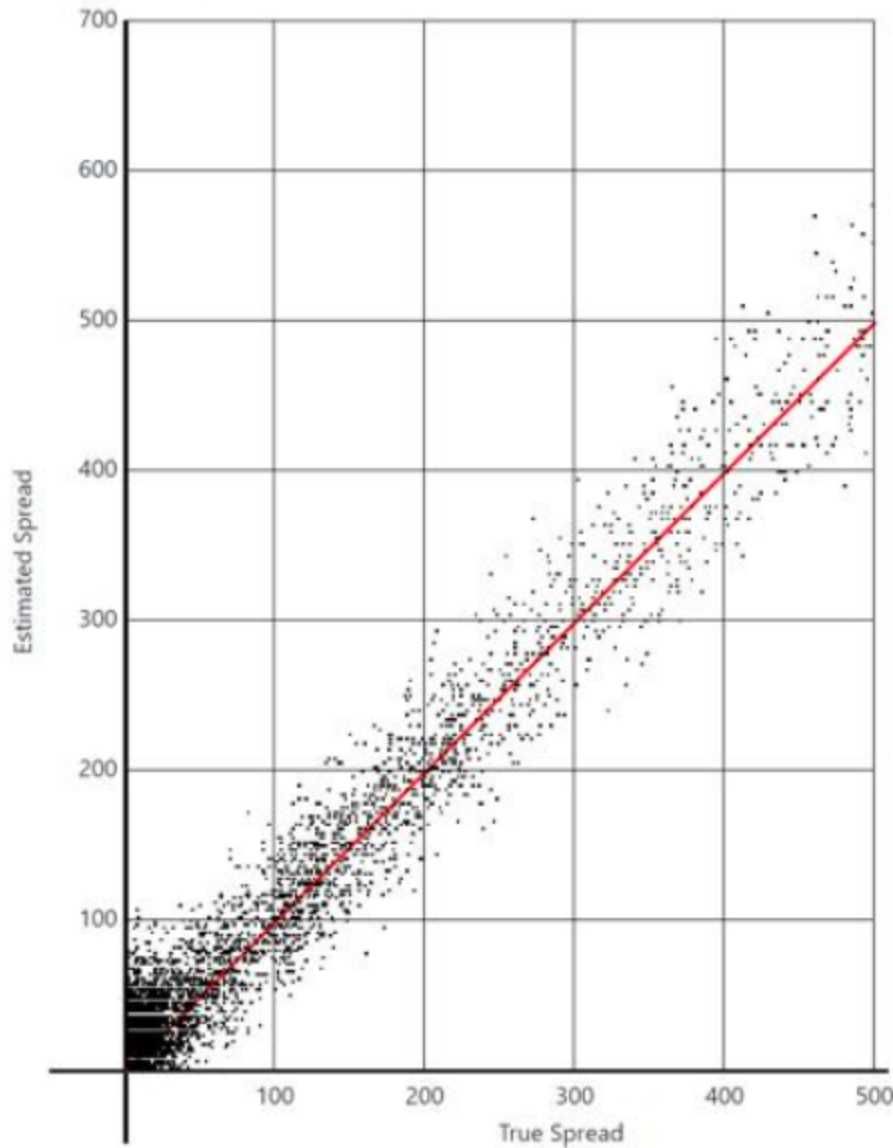
13 7

109 74
129 135
123 157
20 4
2 1
123 84
10 28
106 93
90 111

Output chipped because of large size.

I have also plotted the graph between estimated spread and actual spread.





2. BLOOM SKETCH HYPER LOG LOG:

In this we are maintaining a generic bloom sketch to calculate both spread and size, the data structure hyper log log is maintained via registers and 5 bits in each register. Using geometric hash and computing hashes for the flows we are implementing the formula of calculating flow spread size of Hyper log log.

Source file name is BloomSketchHLL.java where in class constructor i am intializing all input parameters and computing geometric hash and random hash values and computing hash for random flow ids.

The output is being written to a output file called OutputProject5.txt

This is the output being generated for the demo input.

flowID: 23.21.88.60	estimated value: 58375	actual value: 52296
flowID: 85.97.61.130	estimated value: 55344	actual value: 57714
flowID: 159.178.28.17	estimated value: 50361	actual value: 55453
flowID: 122.114.8.30	estimated value: 49815	actual value: 52238
flowID: 36.99.119.77	estimated value: 42939	actual value: 45126
flowID: 107.77.221.58	estimated value: 38065	actual value: 36234
flowID: 67.151.60.134	estimated value: 33608	actual value: 31002
flowID: 189.144.210.6	estimated value: 28369	actual value: 28522
flowID: 213.254.138.15	estimated value: 22903	actual value: 22224
flowID: 183.83.28.209	estimated value: 22708	actual value: 25189
flowID: 40.97.156.5	estimated value: 20499	actual value: 22182
flowID: 205.203.100.47	estimated value: 20145	actual value: 20095
flowID: 181.110.119.34	estimated value: 18867	actual value: 15520
flowID: 115.231.102.115	estimated value: 12433	actual value: 12101
flowID: 106.51.225.144	estimated value: 12166	actual value: 10393
flowID: 52.30.195.21	estimated value: 11499	actual value: 10120
flowID: 52.87.101.173	estimated value: 11435	actual value: 10601
flowID: 88.247.52.231	estimated value: 11197	actual value: 11114
flowID: 202.71.21.94	estimated value: 9825	actual value: 8270

flowID: 183.249.174.157 estimated value: 8183 actual value: 8822
flowID: 106.110.3.233 estimated value: 7198 actual value: 9040
flowID: 117.180.224.203 estimated value: 6981 actual value: 7810
flowID: 106.51.32.3 estimated value: 5765 actual value: 5228
flowID: 199.38.165.28 estimated value: 4363 actual value: 4370
flowID: 52.221.20.79 estimated value: 3683 actual value: 3089