ITU KALEIDOSCOPE
ONLINE 2020

7-11 December 2020

STCCS: Segmented Time Controlled Count Min Sketch

Ismail Khram
Beirut Arab University, Lebanon

Session 8: Security in industrial application

Paper S8.3

ITU KALEIDOSCOPE ONLINE2020

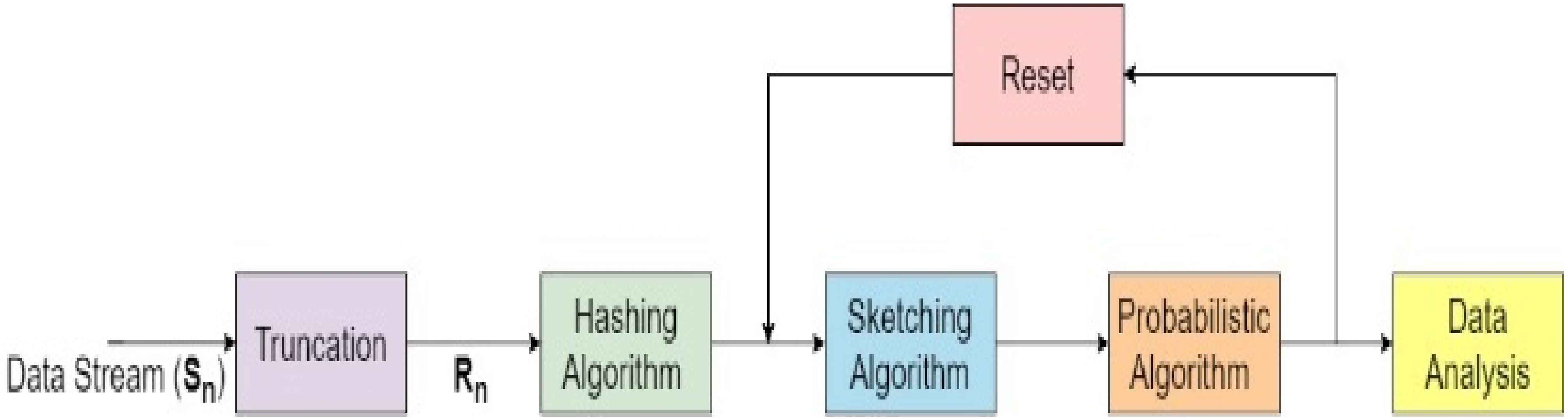
ABSTRACT

- IoT is a concept consisting of many components powered by different techniques and technologies. However, due to computation restrictions, encryption algorithms had to be adapted, often at the expense of lower data security levels and strength. To maintain data privacy between source and sink we present in this paper a data sketching algorithm that utilizes bandwidth by providing a summary of the data to the cloud. The input data stream goes through a hashing algorithm which produces a hexadecimal representation of the data before going through the sketching algorithm. At the algorithm the data is categorized, and the corresponding hash cell value updated. Note is also taken of the arrival time of the data considered anomalous to allow the manager to take corrective action if it is deduced that the periodic appearance of the information is successive in nature.

MOTIVATION & CONTRIBUTIONS

- In the literature care was only taken to detect anomalous data, and providing summaries via sketches
- Sketch variations would only give an impression that the data was anomalous if and when it crosses certain thresholds.
- Sketches would provide false positives ie information about a particular data item might be incorrect.

- Collision reduction with the added hash function depth.
- False positive reduction by segmenting the data into four quadrants. (Normal, Mild, High, Critical)
- Introducing the time element as a way of determining the frequency of the anomalous data. (Close, Far)
- Reset the sketch to avoid saturation.
- Sketch compactness is maintained.


Reset
Data Stream (S_n)
Truncation
R_n
Hashing Algorithm
Sketching Algorithm
Probabilistic Algorithm
Data Analysis

RESULTS

M	1	2	3	4	5	6	7	8	9	10
S_N	38	37	37.5	37	38	37	39	40	40	40
TIME(S)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

Table 1 shows an example data stream with their corresponding arrival times.

Table 2 depicts the hashing algorithm results of four values across five hashing functions.

	H_1	H_2	H_3	H_4	H_5
37	1	3	4	4	2
38	1	2	3	5	2
39	2	4	3	5	3
40	2	4	5	4	3

RESULTS (cont.)

Table 3 shows the final appearance of the table or sketch after applying our sketching algorithm.

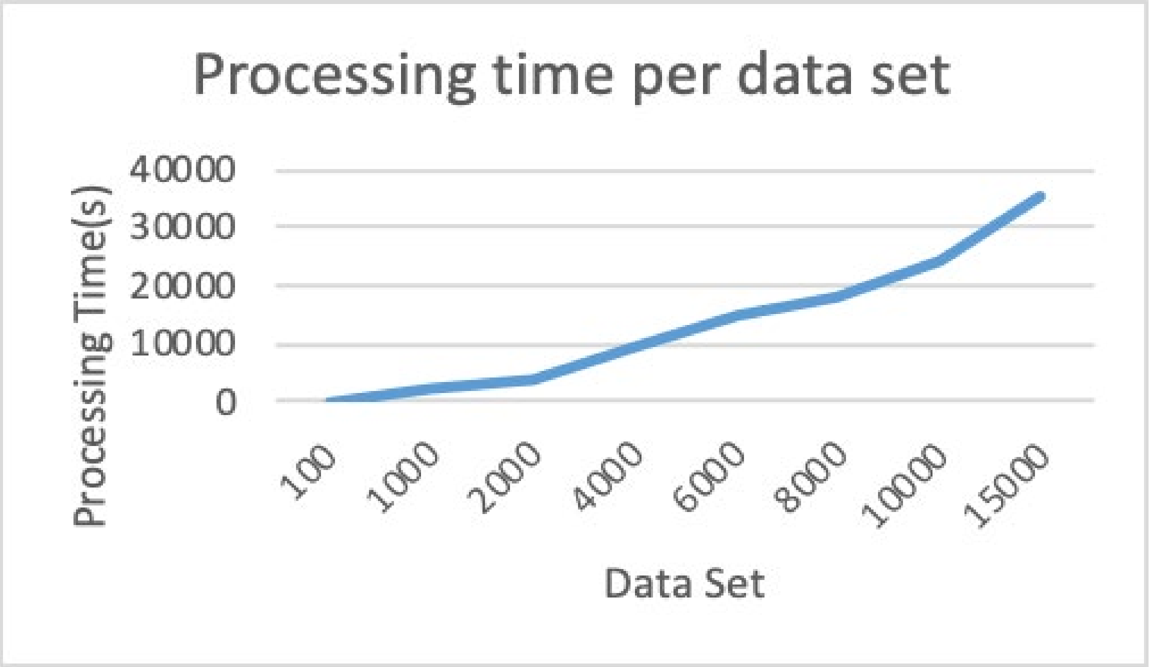
Table 4 shows the change in processing time for an increase in the data set size with the corresponding table size and files size(without table headers)

	NORMAL	MILD	HIGH	CRITICAL	T_1	T_2	T_3	CLOSE	FAR
H_1	4	2	1	3	0.9	1.0	0.6	4	1
H_2	4	2	1	3					
H_3	4	2	1	3					
H_4	4	2	1	3					
H_5	4	2	1	3					

DATA SET	PROCESSING TIME(S)	TABLE SIZE(BYTES)	FILE SIZE(BYTES)
100	208	8388608	8388688
1000	2037	8388608	8388688
2000	4039	8388608	8388688
4000	9351	8388608	8388688
6000	14692	8388608	8388688
8000	17913	8388608	8388688
10000	24407	8388608	8388688
15000	35171	8388608	8388688

RESULTS (cont.)

Figure 1 shows a graphical representation of the variation of processing time with the increase in set size.



CONCLUSION

- Our segmented time-controlled count-min sketch provides a new way of looking at the incoming data. The segmented time-controlled count-min sketch reduces collisions by providing more hash functions and the choice to clear the sketch after either a maximum elapsed time or data item count is reached. Traditionally, the count-min sketch looks at the minimum values produced by the hash functions for a particular value which could produce misleading results especially if the sketch is saturated, so we decided to section the values into categories based on a predetermined range. We also introduced the time element into the sketch as a way of determining the frequency of the anomalous data. We also proved the compactness of the sketch design by showing its consistency with the data increase.

ITU KALEIDOSCOPE
ONLINE 2020

Thank you!