Technical Report

APPENDIX A
MATHEMATICAL ANALYSIS

In this section, we analyze MultiSketch's error bounds for frequency estimation and compare them to the error bounds of Elastic Sketch.

A. Average absolute error (AAE)

The error of MultiSketch mainly comes from upgrading level and light level. When a flow is found in heavy level, it is almost error-free, because the flow must have been upgraded from upgrading level to heavy level very quickly. When a flow is found in upgrading level, it may be overestimated (due to fingerprint collisions) or underestimated (due to probabilistic replacement). When a flow is not found in the first two levels, the results in light level will be returned. At this time, the error bound of the flow is the same as that of CM-Sketch (refer to its original paper [1]). Therefore, we only calculate the error caused by upgrading level.

Assuming that MultiSketch processes a network stream containing N packets, the number of flows is n, let e_i be the i-th flow, and its frequency is f_i. Then the estimated value $\hat{f}_i$ of flow e_i can be written as:

$$\hat{f}_i = f_i - X_i + Y_i \tag{4}$$

Among them, X_i is the decrease caused by the probabilistic replacement, and Y_i is caused by the fingerprint collision. The upper error bound and lower error bound of MultiSketch are determined by Y_i and X_i respectively. We calculate them separately next.

1) Lower error bound: For $E(X_i)$, when a flow $e_j (j \neq i)$ tries to replace e_i in $cell_1$ with a probability of $\frac{1}{f_i}$, $\hat{f}_i$ has a probability of $\frac{1}{f_i}$ to become 0, and a probability of $\frac{f_i-1}{f_i}$ to become f_i. So at this time we have $E(\hat{f}_i) = 0 \times \frac{1}{f_i} + f_i \times \frac{f_i-1}{f_i} = f_i - 1$ and $E(X_i) = E(f_i - \hat{f}_i) = E(f_i) - E(\hat{f}_i) = 1$. This means that the above operation is equivalent to directly reducing the frequency of e_i by 1.

Since there are a total of N packets in the stream and only w buckets are used to hold them, on average $\frac{N}{w}$ packets are inserted per bucket. And since there are B cells in each bucket on average to accommodate B flows, and the frequency of these flows increase sequentially in the bucket ($cell_1.cnt \leq ... \leq cell_B.cnt$), so the maximum number of decays for item e_i occurs in this case: $cell_2 \sim cell_B$ are already occupied by other flows, e_i is always in $cell_1$, and all the flows coming in later are used to decrease e_i. So it follows that:

$$\begin{aligned} E(X_i) &\leq maximum\ \#\ of\ decays \\ &< \frac{N}{w} \times \frac{1}{B+1} \end{aligned} \tag{5}$$

With Markov's inequality, we can get the lower error bound:

$$\begin{aligned} Pr[f_i - \hat{f}_i \geq \epsilon N] &= Pr[X_i - Y_i \geq \epsilon N] \\ &\leq Pr[X_i \geq \epsilon N] \\ &\leq \frac{E(X_i)}{\epsilon N} \\ &< \frac{1}{\epsilon w(B+1)} \end{aligned} \tag{6}$$

Comparison with Elastic Sketch: Elastic Sketch does not underestimate the flow frequency (see the original paper [24] for the specific proof), so MultiSketch is not as good as Elastic Sketch in the lower error bound.

2) Upper error bound: Assuming that the fingerprint length is $\mathcal{F}$. Assuming that e_i is already in $\mathcal{A}_{hash(e_i)}$. we introduce indicator variables $I_{i,j}, (1 \leq i, j \leq n)$ as:

$$I_{i,j} = \begin{cases} 1, & (i \neq j) \wedge (fp(e_i) = fp(e_j)) \wedge \\ & [hash(e_i) = hash(e_j)] \\ 0, & else \end{cases} \tag{7}$$

The last condition in the brackets ensure that e_i and e_j are mapped to the same bucket, and its probability can be written as

$$Pr\{hash(e_i) = hash(e_j)\} = \frac{1}{w} \tag{8}$$

Here we assume $hash(e_i)$ is much larger than w. Then we have:

$$\begin{aligned} E(I_{i,j}) &\leq Pr[fp(e_i) = fp(e_j)] \times \frac{1}{w} \\ &\leq \frac{1}{range(fingerprint)} \times \frac{1}{w} = \frac{1}{w2^{\mathcal{F}}} \end{aligned} \tag{9}$$

Here we assume the length of the fingerprint is $\mathcal{F}$. Then we can derive the expectant of Y_i

$$E(Y_i) = E(\sum_{j=1}^{n} I_{i,j} f_j) \approx \frac{1}{2} \sum_{j=1}^{n} f_j E(I_{i,j}) \leq \frac{N}{w2^{\mathcal{F}}} \tag{10}$$

Similarly, with Markov's inequality, we can get the upper error bound:

$$\begin{aligned} Pr[\hat{f}_i - f_i \geq \epsilon N] &= Pr[Y_i - X_i \geq \epsilon N] \\ &\leq Pr[Y_i \geq \epsilon N] \\ &\leq \frac{1}{\epsilon w 2^{\mathcal{F}}} \end{aligned} \tag{11}$$

Furthermore, assuming that the error probability is δ. If we set the results of Eq. (8) as δ, we can obtain the following time and space complexity in Table I (Notice that the r and w in the first line are the depth and width of the sketch, respectively).

Since many of the algorithms MultiSketch compared do not have corresponding mathematical analysis given

Freq.	r	w	Space	Insert	Query
CM [21]	$\log\frac{1}{\delta}$	$\frac{2}{\epsilon}$	$O(\frac{1}{\epsilon}\log\frac{1}{\delta})$	$O(\log\frac{1}{\delta})$	$O(\log\frac{1}{\delta})$
NI [25]	$\log\frac{1}{\delta}$	$O(\frac{1}{\epsilon^2 p}+\frac{\sqrt{\log\frac{1}{\delta}}}{\epsilon^2 p^{1.5}\sqrt{m}})$	$O(\frac{\log\frac{1}{\delta}}{\epsilon^2 p}+\frac{\log^{1.5}\frac{1}{\delta}}{\epsilon^2 p^{1.5}\sqrt{m}})$	$O(p\log\frac{1}{\delta})$	$O(\log\frac{1}{\delta})$
MV [15]	$\log\frac{1}{\delta}$	$\frac{2}{\epsilon}$	$O(\frac{1}{\epsilon}\log\frac{1}{\delta}\log n)$	$O(\log\frac{1}{\delta})$	$O(\log\frac{1}{\delta})$
Multi	1	$\frac{1}{\delta\epsilon 2^F}$	$O(\frac{1}{\delta\epsilon 2^F})$	$O(1)$	$O(1)$

TABLE I
COMPARISON OF MULTISKETCH WITH STOA.

in their original paper, we only select some of them to make a table as above. Note that some algorithms in the above table have self-contained parameters, so please go to the corresponding original paper if necessary.

Experimental Results: Next, we compare the theoretical error with the real experimental error. We use the CAIDA dataset; the number of packets is 2.49M, the number of flows is 165K, and the frequency of the largest flows is 17K. The experimental results are shown in Figure 10 below. AE is the abbreviation of "average error". It removes the absolute value symbol compared to AAE (average absolute error). Note that $AE = \frac{1}{|\Phi|}\sum_{e_i\in\Phi}(\hat{f}_i - f_i)$, where Φ is the query set (for frequency estimation, this is all flows). The abscissa is the memory, the range is 0.2MB to 2MB, and the interval is 0.2. The black line is the real AE, and the red line is the upper bound of the theoretical AE (i.e., $\frac{N}{w2^F}$ in Eq. (7)). We can get the theoretical AE by substituting the $N = 2.49M$, real w and F into the formula. Note that we did not draw the lower bound because the lower bound obtained in Section A-A1 is very loose. We drew three pictures with fingerprint lengths = 4, 8, and 12 bits, respectively, and it can be found that the upper error bound fits well.

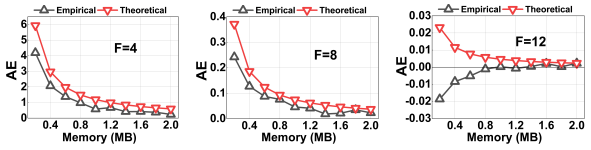


Fig. 10. **Empirical and Theoretical AAE.**

Comparison to Elastic Sketch: We next prove that by adjusting the size of the upgrading level appropriately, the upper error bound of MultiSketch can be better than that of Elastic Sketch.

Assume that the memory of each sketch is M bits, the heavy part of Elastic occupies M_h^e, and the light part (a Count-min Sketch) occupies M_l^e ($M_h^e + M_l^e = M$). Our allocation of the three levels of MultiSketch is as follows: the heavy level also occupies M_h^e, the upgrading level occupies λM_l^e, and the light level occupies $(1-\lambda)M_l^e$, where $0 \le \lambda \le 1$. According to the proof in the original paper of Elastic Sketch [24], if we use an Elastic Sketch to record a stream, the stream can be seen as two sub-streams recorded by the two parts separately. We then have the following definition.

Definition 1. *Let vector* $\boldsymbol{f}=(f_1, f_2, ..., f_n)$ *denote the frequency vector for a stream, where* f_i *denotes the frequency of the i-th flow. Then* f_h *and* f_l *denote the frequency vector of sub-streams recorded by the heavy part and the light part, respectively.*

Now we can give the error bound of Elastic sketch on the frequency estimation task (among them, $X_{i,j}^e$ is a random variable, which represents the hash collision suffered by i-th flow in the j-th layer of the CM Sketch in Elastic Sketch; w and d are the width and depth of CM Sketch):

$$Pr\{\hat{f}_i - f_i \ge \epsilon\|\mathbf{f}_l\|_1\} = Pr\{\forall j, X_{i,j}^e \ge \epsilon\|\mathbf{f}_l\|_1\} \quad (12)$$

$$\le \left(\frac{E(X_{i,j}^e)}{\epsilon\|\mathbf{f}_l\|_1}\right)^d \le \left(\frac{\frac{\|\mathbf{f}_l\|_1}{w}}{\epsilon\|\mathbf{f}_l\|_1}\right)^d \quad (13)$$

$$= \left(\frac{1}{w\epsilon}\right)^d \quad (14)$$

In the original Elastic Sketch paper, the author takes $d = 1$, and the counter size is 8 bits, so we have $\left(\frac{1}{w\epsilon}\right)^d = \frac{8}{M_l^e\epsilon}$.

Next we calculate the error bound of MultiSketch. We assume that the heavy level of MultiSketch is the same as that of Elastic Sketch, records f_h, and is error-free. Then the error of MultiSketch is mainly generated from upgrading and light level. We assume that in the vector f_l, the sub-vector f'_u is the flow frequency recorded by upgrading level, and f'_l is the flow frequency recorded by light level.

For $f_i \in f'_u$, we have (among them, X_i^m is a random variable, which represents the hash and fingerprint collision suffered by flow e_i):

$$Pr\{\hat{f}_i - f_i \ge \epsilon\|\mathbf{f}'_u\|_1\} = Pr\{X_i^m \ge \epsilon\|\mathbf{f}'_u\|_1\} \quad (15)$$

$$\le \frac{E(X_i^m)}{\epsilon\|\mathbf{f}'_u\|_1} \le \frac{\frac{\|\mathbf{f}'_u\|_1}{w_u 2^F}}{\epsilon\|\mathbf{f}'_u\|_1} \quad (16)$$

$$= \frac{1}{w_u 2^F \epsilon} \quad (17)$$

In the experiment, we make each bucket occupy 64 bits, and the fingerprint length is 8 bits, so we have $\frac{1}{w_u 2^F\epsilon} = \frac{1}{4\lambda M_l^e\epsilon}$.

For $f_i \in f'_l$, we have (MultiSketch's light level is the same as Elastic Sketch's.)

$$Pr\{\hat{f}_i - f_i \ge \epsilon\|\mathbf{f}'_l\|_1\} \le \left(\frac{1}{w_m\epsilon}\right)^d \quad (18)$$

Similarly, we take $d = 1$, and the counter size is 8 bits, so we have $\left(\frac{1}{w_m\epsilon}\right)^d = \frac{8}{(1-\lambda)M_l^e\epsilon}$.

In summary, $\forall f_i \in f$, we can derive the upper error bound of MultiSketch (note that $f'_u \cup f'_l = f_l$):

$$Pr\{\hat{f}_i - f_i \geq \epsilon \|\mathbf{f}_l\|_1\} \tag{19}$$

$$= Pr\{f_i \in f'_u\} \cdot Pr\{\hat{f}_i - f_i \geq \epsilon \|\mathbf{f}_l\|_1\} \tag{20}$$

$$+ Pr\{f_i \in f'_l\} \cdot Pr\{\hat{f}_i - f_i \geq \epsilon \|\mathbf{f}_l\|_1\} \tag{21}$$

$$\leq Pr\{f_i \in f'_u\} \cdot Pr\{\hat{f}_i - f_i \geq \epsilon \|\mathbf{f}'_u\|_1\} \tag{22}$$

$$+ Pr\{f_i \in f'_l\} \cdot Pr\{\hat{f}_i - f_i \geq \epsilon \|\mathbf{f}'_l\|_1\} \tag{23}$$

$$= \frac{\|\mathbf{f}'_u\|_1}{\|\mathbf{f}_l\|_1} \cdot \frac{1}{4\lambda M_l^e \epsilon} + \frac{\|\mathbf{f}'_l\|_1}{\|\mathbf{f}_l\|_1} \cdot \frac{8}{(1-\lambda) M_l^e \epsilon} \tag{24}$$

Note that $\|\mathbf{f}'_u\|_1 + \|\mathbf{f}'_l\|_1 = \|\mathbf{f}_l\|_1$. Assuming that $\theta = \frac{\|\mathbf{f}'_u\|_1}{\|\mathbf{f}_l\|_1}$, if we want to make MultiSketch's upper error bound mentioned above smaller than that of Elastic Sketch, we can get the following quadratic inequality:

$$\frac{\theta}{4\lambda M_l^e \epsilon} + \frac{8(1-\theta)}{(1-\lambda) M_l^e \epsilon} < \frac{8}{M_l^e \epsilon} \tag{25}$$

Its solution is $\frac{-b-\sqrt{b^2-16\theta}}{8} < \lambda < \frac{-b+\sqrt{b^2-16\theta}}{8}$, where $b = 33\theta - 28$. The λ in this range can make the upper error bound of MultiSketch better than that of Elastic Sketch.

Appendix B
Detailed Experiment Setup

The entry sizes in CM and AS are 16 or 32 bits, depending on the maximum item size in datasets. CM and CU allocate 4 arrays and use 4 32-bit Bob hash [36] functions for items mapping. AS consists of the widely used CM sketch and a filter. The filter will allocate about 0.4KB of additional memory, and the CM sketch of AS also includes 4 arrays and 4 32-bit Bob hash functions. In the sketches using the SEAD Counter and SAC counter, the size of the counters is reduced to 16 bits. All entries of the PCU are 4 bits, and the number of mapped entries is 4. The PCU uses one 64-bit Bob hash function. EL's heavy part contains 8 entries in each bucket, and the depth of the CM sketch in the light part is 1. The depth of the count sketch of NI is 4, and the geometric sampling rate is $p = 0.01$ (recommended value). The initial counter size of SALSA is 4 bits and the maximum is 32 bits. The bucket size of our MultiSketch is 64 bits, and the cell layout is shown in Figure 11. The heavy level, upgrading level, and elastic level of MultiSketch account for 5%, 85%, and 10% of the total memory, respectively, and the parameter tuning process is omitted.

Appendix C
The Result of Campus Dataset

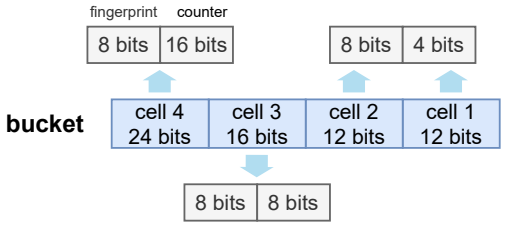
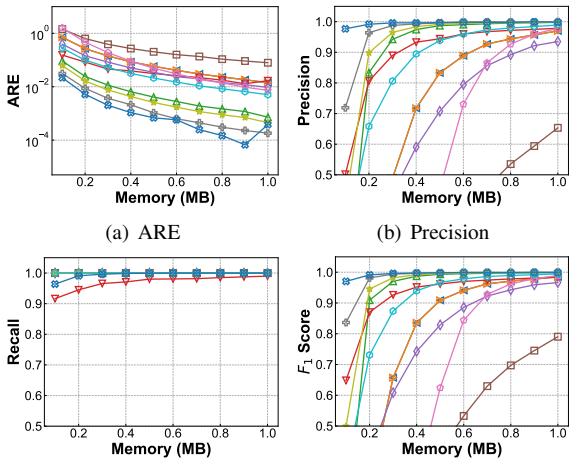


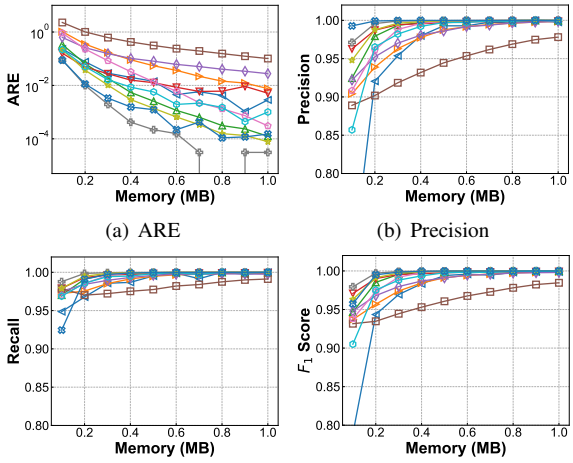
Fig. 11. The specific structure of cells in each 64-bit bucket.



(a) ARE (b) Precision

(c) Recall (d) F_1 Score

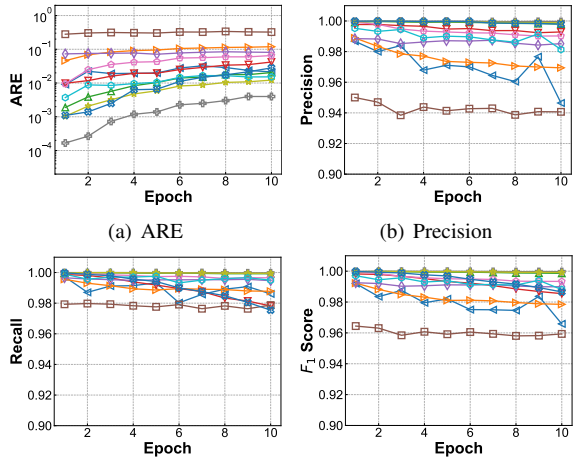
Fig. 12. Heavy hitter detection vs. memory ($0.1 \sim 1.0MB$) - Campus datasets.



(a) ARE (b) Precision

(c) Recall (d) F_1 Score

Fig. 13. Heavy change detection vs. **memory** ($0.1 \sim 1.0MB$) - Campus datasets.



(a) ARE (b) Precision

(c) Recall (d) F_1 Score

Fig. 14. Heavy change detection vs. **epoch** ($1 \sim 10$) - Campus datasets.