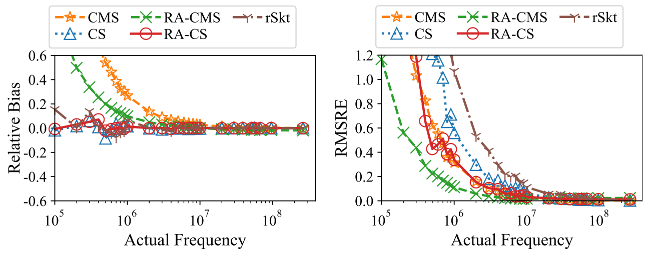
A ADDITIONAL EXPERIMENTAL RESULTS

This section shows the additional experimental results that are not included in the submitted version of this paper.

A.1 Frequency Estimation

In this subsection, we evaluate the performance of our RA-CS in frequency estimation using the CAIDA dataset where each data stream tuple is represented as $(flow\ ID, 1/{-1}\cdot packet\ size\ in\ bytes)$. We set the D:I ratio to 0.5. The metrics used include RB and RMSRE, and the results are shown in Fig. 13(a) and (b), respectively.

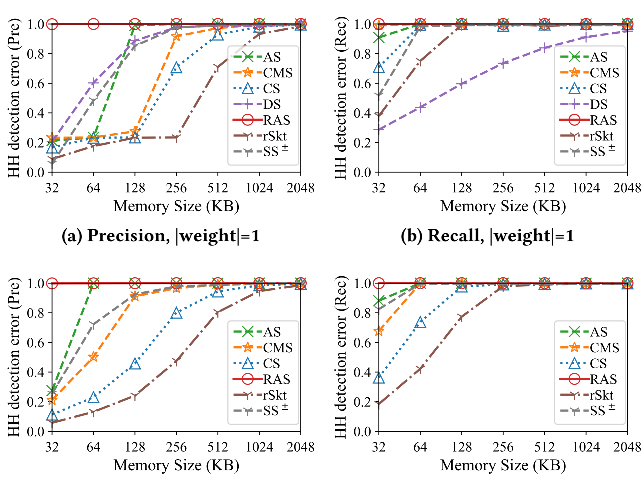


(a) RB, 128 KB of memory **(b) RMSRE, 128 KB of memory**

Figure 13: Frequency estimation errors when |weight| > 1

A.2 Heavy Hitter Detection

In this subsection, we evaluate the performance of RAS in heavy hitter detection using the Precision and Recall metrics. For comparison, we set the relative threshold ε defined in Eq. (2) to 2^{-11} and set the filter size to $\frac{\zeta}{12\varepsilon}$. Firstly, we fix the D:I ratio to 0.5 and use the CAIDA dataset for evaluation. The results for the stream tuple represented as $(flow\ ID, 1/{-1})$ are shown in Fig. 14(a) and (b), and the results for the stream tuple represented as $(flow\ ID, 1/{-1}\cdot packet\ size\ in\ bytes)$ are shown in Fig. 14(c) and (d), respectively.

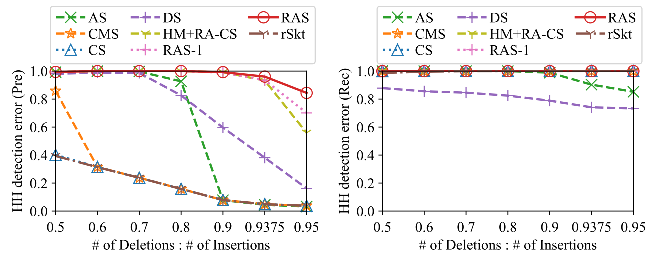


(a) Precision, |weight|=1 **(b) Recall, |weight|=1**

(c) Precision, |weight|>1 **(d) Recall, |weight|>1**

Figure 14: Precision and Recall for heavy hitter detection when D:I ratio is 0.5

Then, we evaluate the performance of our RAS under different D:I ratios using the Zipf distribution dataset. The results for the two metrics are shown in Fig. 15(a) and (b), respectively.

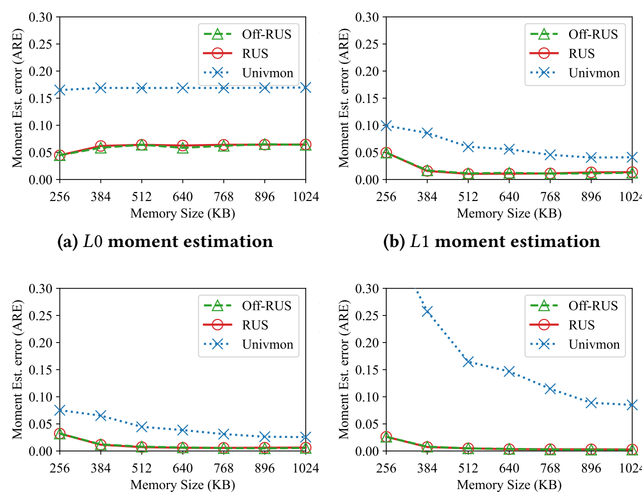


(a) Precision **(b) Recall**

Figure 15: Precision and Recall for heavy hitter detection under varying D:I ratios

A.3 Moment Estimation

First, we evaluate the performance of our algorithm in moment estimation using the CAIDA dataset when the stream tuples are in the form of $(flow\ ID, 1/{-1})$. We set the D:I ratio to 0.5, and use the ARE as a metric to evaluate the estimation performance. The estimation error curves of the $L0$ moment, $L1$ moment, entropy, and $L2$ moment, are shown in Fig. 16(a), (b), (c), and (d), respectively.

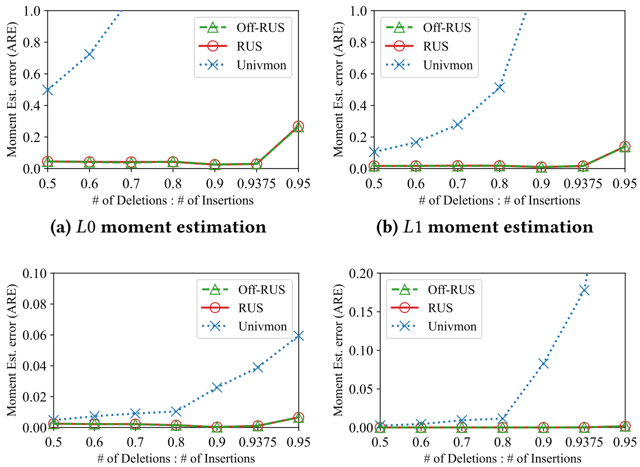


(a) $L0$ moment estimation **(b) $L1$ moment estimation**

(c) Entropy estimation **(d) $L2$ moment estimation**

Figure 16: Moment estimation errors when |weight| = 1

Then, we evaluate the performance of our algorithm in moment estimation under varying D:I ratios. We use the synthetic Zipf distribution dataset. The estimation error curves of the $L0$ moment, $L1$ moment, entropy, and $L2$ moment, are shown in Fig. 17(a), (b), (c), and (d), respectively.



(a) $L0$ moment estimation
(b) $L1$ moment estimation
(c) Entropy estimation
(d) $L2$ moment estimation

Figure 17: Moment estimation errors on varying D:I ratios

B MATHEMATICAL ANALYSIS

In this section, we present the mathematical analyses of our solutions. We first analyze the expectation and variance of our RAC. Then, we derive the L1 and L2 error bounds of our RA-CS. Finally, for RAS, we provide the lower bound of the filter size and the error bound of the heavy hitter detection error.

B.1 Mathematical Analysis of RAC

In this subsection, we provide the expectation of RAC which proves that our RAC is unbiased. Then we give the variance of our RAC, and the coefficient of variation (i.e., $\frac{variance}{expectation^2}$).

Suppose an RAC C with an exponent α of length $\mathcal{L}_\alpha$, and a coefficient β of length $\mathcal{L}_\beta$. Denote by $T(a,b)$ the random variable that represents the amount of total update value needed from the start of counting ($C.\alpha = 0, C.\beta = 0$) until that RAC C reaches the state when $C.\alpha$ equals a and $C.\beta$ equals b. We also assume a uniform increment or decrement size θ.

Theorem 2. *Suppose that increments to the RAC C are uniform and given by θ. Then, for the random variable $T(a,b)$ defined above we have:*

$$E[T(a,b)] = (b + 2^{\mathcal{L}_\beta}) \cdot 2^a - 2^{\mathcal{L}_\beta}$$

Proof. Recall that for a geometric random variable $G(p)$, we have $E[G(p)] = \frac{1}{p}$ and $Var[G(p)] = \frac{1-p}{p^2}$. For each a, denote by $W(a)$ the random variable that represents the amount of value to make one increment of the RAC C when $C.\alpha = a$. The probability of incrementing the RAC C in a single trial is $p_a = \frac{\theta}{2^a}$. Note that the maximum value of $C.\alpha$ is $2^{\mathcal{L}_\alpha} - 1$. Therefore, the number of trials before the increment of RAC C is $G(p_a)$ and since each trial corresponds to θ of the total frequency we have

$$W(a) = \theta G(p_a)$$

Time from the beginning of counting can be divided in $2^{\mathcal{L}_\alpha}$ intervals, corresponding to each exponent $\alpha = 0, 1, ..., a < 2^{\mathcal{L}_\alpha}$. In exponent $\alpha = 0$, RAC C is incremented $q_0 = 2^{\mathcal{L}_\beta}$ times, for exponent $\alpha = 1, 2, ..., a \leq 2^{\mathcal{L}_\alpha} - 2$, we have $q_a = 2^{\mathcal{L}_\beta}$ times. For exponent $\alpha = 2^{\mathcal{L}_\alpha} - 1$, we have $q_a = 2^{\mathcal{L}_\beta} - 1$. This is because if we increment $2^{\mathcal{L}_\beta}$ times, the coefficient will overflow, and the exponent will increase 1 to correct this overflow, however, the maximum value of exponent α is $2^{\mathcal{L}_\alpha} - 1$. If $a = 2^{\mathcal{L}_\alpha}$, it overflows.

Denote by $Q(a)$ the total increments in each of these $2^{\mathcal{L}_\alpha}$ intervals. Now for any $\alpha = 0, 1, ..., a$.

$$Q(a) = \sum_{j=1}^{q_a} W_j(a)$$

where $W_j(a)$ is the set of independent identically distributed (i.i.d.) random variables, with distribution given by $W(a)$.

Therefore

$$E[Q(a)] = \sum_{j=1}^{q_a} E[W_j(a)] = q_a \cdot \theta \frac{1}{p_a} = q_a \cdot 2^a$$

where $a \in \{0, 1, \ldots, 2^{\mathcal{L}_\alpha} - 1\}$, and $p_a = \frac{\theta}{2^a}$.

Now $T(a,b)$ is given by:

$$T(a,b) = \sum_{\alpha=0}^{a} Q(\alpha)$$

When a is 0, the expected value of $T(a,b)$ is

$$\begin{aligned} E[T(a,b)] &= E[\sum_{j=1}^{b} W_j(0)] \\ &= \sum_{j=1}^{b} E[\theta G(p_0)] \\ &= b \end{aligned}$$

The theorem follows when $a = 0$.

When $a \in 1, 2, \ldots, 2^{\mathcal{L}_\alpha} - 2$, the expected value of $T(a,b)$ is

$$\begin{aligned} E[T(a,b)] &= E[Q(0)] + \sum_{\alpha=1}^{a-1} E[Q(\alpha)] + b \cdot E[W(a)] \\ &= 2^{\mathcal{L}_\beta} + 2^{a+\mathcal{L}_\beta} - 2^{\mathcal{L}_\beta+1} + b \cdot 2^a \\ &= (b + 2^{\mathcal{L}_\beta}) \cdot 2^a - 2^{\mathcal{L}_\beta} \end{aligned}$$

The theorem follows when $a \in 1, 2, \ldots, 2^{\mathcal{L}_\alpha} - 2$.

When $a = 2^{\mathcal{L}_\alpha} - 1$, the expected value of $T(a,b)$ is

$$\begin{aligned} E[T(a,b)] &= E[Q(0)] + \sum_{\alpha=1}^{2^{\mathcal{L}_\alpha}-2} E[Q(\alpha)] + b \cdot E[W(a)] \\ &= 2^{\mathcal{L}_\beta} + 2^{2^{\mathcal{L}_\alpha}+\mathcal{L}_\beta-1} - 2^{\mathcal{L}_\beta+1} + b \cdot 2^{2^{\mathcal{L}_\alpha}-1} \\ &= (b + 2^{\mathcal{L}_\beta}) \cdot 2^a - 2^{\mathcal{L}_\beta} \end{aligned}$$

The theorem follows when $a = 2^{\mathcal{L}_\alpha} - 1$. □

Corollary 3. *The estimation of RAC C is unbiased.*

Proof. As stated in Theorem 2, the amount of total update value is equal to the expectation of the estimation of RAC C. □

Theorem 4. *Suppose that increments to the RAC C are uniform and given by θ. Then, for the random variable $T(a,b)$ defined above we have:*

$$Var[T(a,b)] \leq \frac{1}{3}\left(2^{2^{\mathcal{L}_\alpha}}-2\right)2^{\mathcal{L}_\beta}\left(-3\theta+2^{2^{\mathcal{L}_\alpha}}+2\right) - 2^{2^{\mathcal{L}_\alpha}-2}\left(2^{2^{\mathcal{L}_\alpha}}-2\theta\right)$$

Proof. Since $Q(a)=\sum_{j=1}^{q_a}W_j(a)$, we have

$$\begin{aligned}Var[Q(a)] &= \sum_{j=1}^{q_a} Var[W_j(a)]\\ &= \sum_{j=1}^{q_a} Var[\theta G(p_a)]\\ &= q_a\cdot 2^a\cdot 2^a\cdot(1-p_a)\end{aligned}$$

The variance of $T(a,b)$ is now:

$$\begin{aligned}Var(T(a,b)) &= \sum_{\alpha=0}^{a} Var[Q(a)]\\ &\leq Var[Q(0)] + \sum_{a=1}^{2^{\mathcal{L}_\alpha}-2} Var[Q(a)] + Var[Q(\mathcal{L}_\alpha-1)]\end{aligned}$$

For $Var[Q(0)]$, as the exponent part is 0, the increment is deterministic, so the variance is 0.

For $\sum_{\alpha=1}^{\mathcal{L}_\alpha-2} Var[Q(\alpha)]$, we have:

$$\begin{aligned}\sum_{\alpha=1}^{2^{\mathcal{L}_\alpha}-2} Var[Q(\alpha)] &= \sum_{\alpha=1}^{2^{\mathcal{L}_\alpha}-2} q_\alpha\cdot 2^{2\alpha}(1-p_\alpha)\\ &= \sum_{\alpha=1}^{2^{\mathcal{L}_\alpha}-2} 2^{\mathcal{L}_\beta}\cdot 2^{2\alpha}\left(1-\frac{\theta}{2^\alpha}\right)\\ &= \frac{1}{3}2^{2^{\mathcal{L}_\alpha+1}+\mathcal{L}_\beta-2} - \frac{1}{3}2^{\mathcal{L}_\beta+2} - \theta 2^{\mathcal{L}_\beta+2^{\mathcal{L}_\alpha}-1} + \theta 2^{\mathcal{L}_\beta+1}\end{aligned}$$

For $Var[Q(2^{\mathcal{L}_\alpha}-1)]$, we have:

$$\begin{aligned}Var[Q(2^{\mathcal{L}_\alpha}-1)] &= Var\left[\sum_{j=1}^{q_{2^{\mathcal{L}_\alpha}-1}} W_j(2^{\mathcal{L}_\alpha}-1)\right]\\ &= (2^{\mathcal{L}_\beta}-1)\cdot(2^{2^{\mathcal{L}_\alpha+1}-2}-\theta 2^{2^{\mathcal{L}_\alpha}-1})\end{aligned}$$

Overall, the Variance:

$$\begin{aligned}Var[T(a,b)] &\leq \sum_{a=0}^{\mathcal{L}_\alpha-1} Var[Q(a)]\\ &= \frac{1}{3}\left(2^{2^{\mathcal{L}_\alpha}}-2\right)2^{\mathcal{L}_\beta}\left(-3\theta+2^{2^{\mathcal{L}_\alpha}}+2\right)\\ &\quad -2^{2^{\mathcal{L}_\alpha}-2}\left(2^{2^{\mathcal{L}_\alpha}}-2\theta\right)\end{aligned}$$

□

The following corollary characterizes the asymptotic behavior of the coefficient of variation $\delta(T(A,m))$.

Corollary 5. *For RAC C, $\delta(T(a,b)) \approx \sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}}$.*

Proof.

$$\delta(T(a,b)) = \sqrt{\frac{Var[T(a,b)]}{(E[T(a,b)])^2}}$$

And we have

$$\begin{aligned}\frac{Var[T(a,b)]}{(E[T(a,b)])^2} &= \frac{\left(2^{2^{\mathcal{L}_\alpha}}-2\right)2^{\mathcal{L}_\beta+2}\left(-3\theta+2^{2^{\mathcal{L}_\alpha}}+2\right)-3\,2^{2^{\mathcal{L}_\alpha}}\left(2^{2^{\mathcal{L}_\alpha}}-2\theta\right)}{3\left(2^{2^{\mathcal{L}_\alpha}}-\left(2^{2^{\mathcal{L}_\alpha}}-1\right)2^{\mathcal{L}_\beta+1}\right)^2}\\ &\approx \frac{2^{\mathcal{L}_\beta+2}-3}{3\left(2^{\mathcal{L}_\beta+1}-1\right)^2}\\ &\approx \frac{2^{-\mathcal{L}_\beta}}{3}\end{aligned}$$

where the approx symbol follows from that the $2^{2^{\mathcal{L}_\alpha}}$ is larger than the constant term in the addition process. □

B.2 Mathematical Analysis of RA-CS

In this subsection, we first prove that the estimation of RA-CS (i.e., CountSketch combined with our RAC) is unbiased. Then, we provide the L1 and L2 error bound of the RA-CS.

Theorem 6. *The RA-CS provides an unbiased element frequency estimation.*

Proof. We treat the value of RAC C as the true value C_r plus the estimation error C_ϵ, such that $C=C_r+C_\epsilon$. For estimating the frequency of an element e, in each row of the CountSketch, we have

$$\hat{f_e} = f_e + \sum_{e':e'\neq e} f_{e'}g(e)g(e')Y_{e'}$$

where

$$Y_{e'} = \begin{cases}1, & if\ \ h(e')=h(e)\\ 0, & otherwise\end{cases}$$

So, for the frequency estimation $\hat{f_e}$ of the element e, we have

$$\begin{aligned}E[\hat{f_e}] &= E[C]\\ &= E[C_r]+E[C_\epsilon]\\ &= f_e + E\left[\sum_{e':e'\neq e} f_{e'}g(e)g(e')Y_{e'}\right] + E[C_\epsilon]\end{aligned}$$

Since the CountSketch and RAC C are unbiased, we have $E[C_r]=0$ and $E[\sum_{e':e'\neq e} f_{e'}g(e)g(e')Y_{e'}]=0$. Therefore, $E[\hat{f_e}]=f_e$, and the theorem follows. □

Theorem 7 (L1 error bound for CountSketch with general depth). *For a RA-CS with depth d, and width b, the frequency estimation error $|f_e-\hat{f_e}|$ of an element e is at most $\frac{\epsilon}{b}F_1$ with probability $1-\delta$ from the RAS with parameters $d=O(\log(1/\delta))$.*

Proof. For the L1 bound in one row, we have

$$\begin{aligned}E[|f_e-\hat{f_e}|] &= E\left[\left|\sum_{e'\neq e} Y_{e'}\cdot f_{e'}\cdot g_j(e)\cdot g_j(e') + C_\epsilon\right|\right]\\ &\leq E\left[\left|\sum_{e'\neq e} Y_{e'}\cdot f_{e'}\cdot g_j(e)\cdot g_j(e')\right|\right] + E[|C_\epsilon|]\\ &= E\left[\left|\sum_{e'\neq e} Y_{e'}\cdot f_{e'}\cdot g_j(e)\cdot g_j(e')\right|\right] + E[|C-C_r|]\end{aligned}$$

For $E[|\sum_{e'\neq e} Y_{e'}\cdot f_{e'}\cdot g_j(e)\cdot g_j(e')|]$, we have

$$E\left[\left|\sum_{e'\neq e} Y_{e'}\cdot f_{e'}\cdot g_j(e)\cdot g_j(e')\right|\right] \leq \frac{F_1}{b}$$

for $E[|C-C_r|]$, since the RAC C is unbiased (i.e., $C_r = E[C]$), we have

$$
\begin{aligned}
E[|C-C_r|] &\le \sqrt{E[(C-C_r)^2]} = \sqrt{Var[C]} \\
&= \sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}\cdot(E[C])^2} \\
&= C_r \cdot \sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}}
\end{aligned}
$$

So, we have

$$E[|f_e - \hat{f_e}|] \le \frac{F_1}{b} + C_r \cdot \sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}}$$

Where the C_r can be seen as the expectation of the frequency in one RAC, which is $\frac{F_1}{b}$.

So

$$
\begin{aligned}
E[|f_e - \hat{f_e}|] &\le \frac{F_1}{b} + \frac{F_1}{b}\cdot \sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}} \\
&= \frac{F_1}{b}\cdot(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})
\end{aligned}
$$

Note that the b equals twice that in the vanilla CountSketch.

Combine Markov's inequality, we have

$$P[|f_e - \hat{f_e}| \ge \frac{\epsilon}{b}F_1] \le \frac{1}{\epsilon}\cdot(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})$$

Setting ϵ to 3, we have

$$P[|f_e - \hat{f_e}| \ge \frac{3}{b}F_1] \le \frac{1}{3}\cdot(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})$$

$$P[|f_e - \hat{f_e}| \le \frac{3}{b}F_1] \ge 1-\frac{1}{3}\cdot(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})$$

Combining the Chernoff bounds, Theorem 7 follows. □

Theorem 7 suggests using a value of b that is logarithmic in the desired failure probability. However, practitioners rarely use more than a small constant number of rows, such as 3, 4, or 5 rows. Recently, a study [29] proved that a CountSketch with a depth of 3 satisfies a similar error bound:

Theorem 8 (L1 error bound for CountSketch with a depth of 3). *For a RA-CS with a depth of* 3*, and width of* b*, the frequency estimation error* $|f_e - \hat{f_e}|$ *of an element* e *is at most* $\frac{\sqrt{3}}{b}(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})F_1$.

Proof. Combining Corollary 5 with the study [29], the theorem follows. □

Theorem 9 (L2 error bound for CountSketch with general depth). *For a RAS with depth* d*, and width* b*, the frequency estimation error* $|f_e - \hat{f_e}|$ *of an element* e *is at most* $\frac{\epsilon}{\sqrt{b}}F_2$ *with probability* $1-\delta$ *from the RAS with parameters* $d = O(\log(1/\delta))$.

Proof. We have computed the variance of the RAC C, based on the expectation $E[C] = C_r$. So we have the conditional variance $Var[C|C_r]$. So we have

$$Var[C] = E[Var[C|C_r]] + Var[E[C|C_r]]$$

where $Var[C|C_r] = Var[T(a,b)] \approx \frac{2^{-\mathcal{L}_\beta}}{3}\cdot(E[T(a,b)])^2 < \frac{2^{-\mathcal{L}_\beta}}{3}\cdot\frac{{F_2}^2}{b}$, and $Var[E[C|C_r]]$ is $Var[C_r] \le \frac{{F_2}^2}{b}$.

$$
\begin{aligned}
Var[C] &\le E[\frac{2^{-\mathcal{L}_\beta}}{3}\cdot(E[T(a,b)])^2] + \frac{{F_2}^2}{b} \\
&= \frac{2^{-\mathcal{L}_\beta}}{3}E[(E[T(a,b)])^2] + \frac{{F_2}^2}{b} \\
&\le \frac{2^{-\mathcal{L}_\beta}}{3}\frac{{F_2}^2}{b} + \frac{{F_2}^2}{b} \\
&= \frac{3+2^{-\mathcal{L}_\beta}}{3}\frac{{F_2}^2}{b}
\end{aligned}
$$

Note that the b is twice as large as that of the CountSketch.

By the Chebyshev's inequality, we have

$$P[|f_e - \hat{f_e}| \ge \frac{\epsilon}{\sqrt{b}}F_2] \le \frac{3+2^{-\mathcal{L}_\beta}}{3}\cdot\frac{1}{\epsilon^2}$$

Setting ϵ to 2, we have

$$P[|f_e - \hat{f_e}| \ge \frac{2}{\sqrt{b}}F_2] \le \frac{3+2^{-\mathcal{L}_\beta}}{3}\cdot\frac{1}{4}$$

$$P[|f_e - \hat{f_e}| \le \frac{2}{\sqrt{b}}F_2] \ge 1-\frac{3+2^{-\mathcal{L}_\beta}}{3}\cdot\frac{1}{4}$$

By the Chernoff bounds, Theorem 9 follows. □

Theorem 10 (L2 error bound for CountSketch with a depth of 3). *For a RAS with depth* 3*, and width* b*, the frequency estimation error* $|f_e - \hat{f_e}|$ *of an element* e *is at most* $\sqrt{\frac{3+2^{-\mathcal{L}_\beta}}{3}}\frac{F_2}{\sqrt{b}}$.

Proof. Combining Corollary 5 with the study [29], the theorem follows. □

B.3 Mathematical Analysis of RAS

In this subsection, we first present the lower bound on the size of the prefilter. Then, we propose the error bound of the frequency estimation of the heavy hitters.

Lower Bound on Filter Size. Previously, the SpaceSaving$^\pm$ (SS$^\pm$) [64] established a lower bound $L = \frac{\zeta}{\varepsilon}$ on its size to track all HHs in the *bounded deletion model*. However, the lower bound is determined under the assumption of an evenly distributed insertion pattern. In this pattern, the number of insertions of each element is equal; therefore, no HHs exist. In this section, we further derive the lower bound. We consider a data stream following Zipf Law, which is characterized by heavy hitters, and most real-world data streams exhibit this distribution.

The lower bound when the data streams follow a Zipf distribution is given in Theorem 1, whose proof is as follows.

Proof. The probability mass function of the Zipf distribution is shown in Eq. (11), where η represents the exponent that characterizes the skewness of the distribution, and N is the number of distinct elements.

$$f(i;\eta,N) = 1\ /\ \left(i^{\eta}\sum\nolimits_{j=1}^{N} j^{-\eta}\right) \tag{11}$$

After all I insertions and D deletions, the frequencies of heavy hitters are at least $\frac{\varepsilon}{\zeta}I$, since $\varepsilon(I-D)=\frac{\varepsilon}{\zeta}I$. Thus, the prefilter must retain all the elements with a frequency greater than $\frac{\varepsilon}{\zeta}I$ before any deletion occurs. Otherwise, if such an element is not retained before deletion occurs, it becomes an untracked heavy hitter afterward. Therefore, the size lower bound L is the maximum value of k that satisfies $f(k;\eta,N)\geq\frac{\varepsilon}{\zeta}$. The expression of L is defined as

$$L=\max k,\ \text{s.t. } 1\leq k\leq N\ \wedge\ f(k;\eta,N)\geq\frac{\varepsilon}{\zeta}.$$

Combining Eq. (11), the value of k can be determined as

$$k\leq\left\lfloor\sqrt[\eta]{\frac{\zeta/\varepsilon}{\sum_{i=1}^{N}(1/i)^{\eta}}}\right\rfloor.$$

□

The derivative of $\sqrt[\eta]{\frac{\zeta/\varepsilon}{\sum_{i=1}^{N}(1/i)^{\eta}}}$ for η is negative, i.e., $\frac{\partial}{\partial\eta}\sqrt[\eta]{\frac{\zeta/\varepsilon}{\sum_{i=1}^{N}(1/i)^{\eta}}}<0$. As the parameter η describes the skewness, this means that a Zipf distribution with less skewness needs a higher lower bound on the size of the prefilter. Consequently, the configuration of k works for a low-skew Zipf distribution and is also compatible with the high-skew Zipf distribution. So we derive the corollary of the lower bound k when $\eta=1$ (i.e., low skew Zipf distribution) to fit more scenarios.

Corollary 11. *When the skewness η of a Zipf distribution is* 1, *the lower bound can be set to* $\frac{\zeta}{12\varepsilon}$.

Proof. When η is 1, the size lower bound L can be written as

$$k=\frac{\zeta/\varepsilon}{\sum_{i=1}^{N}(1/i)}\approx\frac{\zeta/\varepsilon}{ln(N+1)+\gamma}, \tag{12}$$

where $\gamma\approx 0.5772$ is the Euler's constant.

By combining Theorem 1 with the assumption that N typically exceeds 2^{16} in real-world datasets, we set the size of the filter to $\frac{\zeta/\varepsilon}{12}$, which is only $\frac{1}{12}$ of that of the SS$^{\pm}$. This value is validated to yield excellent results in the evaluation presented in Section 9. □

Estimation Error of RAS. From Theorem 1, we transform the ε-heavy hitters problem to the top-k heavy hitter problem, where k is set to $\frac{\zeta}{\epsilon\sum_{i=1}^{N}(1/i)}$. Let f_e^k denote the frequency of the kth frequent element. We then give the L1 and L2 error bound to ensure that all the elements with frequency at least $(1-\epsilon)f_e^k$ are maintained.

Theorem 12 (L1 bound of the heavy hitter problem with a depth of 3). *If d is set to* 3, *and* $b\geq\frac{2\sqrt{3}(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})F_1}{f_e^k}$, *then the element with frequency no less than* $(1-\epsilon)f_e^k$ *in top-*$\frac{\zeta}{\epsilon\sum_{i=1}^{N}(1/i)}$ *element are preserved.*

Proof. By Theorem 7, the estimation for the frequency of all elements is within an additive factor of $\frac{\sqrt{3}}{b}(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})F_1$ of the actual element frequency. Thus for two elements whose true frequency differs by more than $2\cdot\frac{\sqrt{3}}{b}(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})F_1$, the estimation can correctly identify the more frequent element. By setting $2\cdot\frac{\sqrt{3}}{b}(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})F_1\leq\epsilon f_e^k$, we ensure that the only elements that can replace the true most frequent elements in the estimated top-$\frac{\zeta}{\epsilon\sum_{i=1}^{N}(1/i)}$ are elements with true frequency at least $(1-\epsilon)f_e^k$.

$$2\cdot\frac{\sqrt{3}}{b}(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})F_1\leq\epsilon f_e^k$$
$$b\geq\frac{2\sqrt{3}(1+\sqrt{\frac{2^{-\mathcal{L}_\beta}}{3}})F_1}{f_e^k}$$

□

Theorem 13 (L1 bound of the heavy hitter problem with general depth). *If* $b\geq\frac{\epsilon f_e^k}{6F_1}$, *then the element with frequency no less than at least* $(1-\epsilon)f_e^k$ *in top-*$\frac{\zeta}{\epsilon\sum_{i=1}^{N}(1/i)}$ *element are preserved with probability* $1-\delta$ *from the RAS with parameters* $d=O(\log(1/\delta))$.

Proof. Similar to Theorem 12, we set the error smaller than half of ϵf_e^k, and the theorem follows. □

Theorem 14 (L2 bound of the heavy hitter problem with a depth of 3). *If d is set to* 3, *and* $b\geq\frac{4}{9}{F_2}^2\frac{(3+2^{-\mathcal{L}_\beta})^2}{(\epsilon f_e^k)^2}$, *then the element with frequency no less than at least* $(1-\epsilon)f_e^k$ *in top-*$\frac{\zeta}{\epsilon\sum_{i=1}^{N}(1/i)}$ *element are preserved.*

Proof. The proof is similar to that of Theorem 12. Combined with Theorem 10, when $b\geq\frac{4}{9}{F_2}^2\frac{(3+2^{-\mathcal{L}_\beta})^2}{(\epsilon f_e^k)^2}$, the estimation error is smaller than half of $(1-\epsilon)f_e^k$, so the theorem follows. □

Theorem 15 (L2 bound of the heavy hitter problem with general depth). *If* $b\geq\frac{16{F_2}^2}{\epsilon^2{f_e^k}^2}$, *then the element with frequency no less than at least* $(1-\epsilon)f_e^k$ *in top-*$\frac{\zeta}{\epsilon\sum_{i=1}^{N}(1/i)}$ *element are preserved with probability* $1-\delta$ *from the RAS with parameters* $d=O(\log(1/\delta))$.

Proof. The proof is similar to that of Theorem 12. Combined with Theorem 9, when $b\geq\frac{16{F_2}^2}{\epsilon^2{f_e^k}^2}$, the estimation error is smaller than half of $(1-\epsilon)f_e^k$, so the theorem follows. □

REFERENCES

[1] [n.d.]. A Universal Sketch for Estimating Heavy Hitters and Per-Element Frequency Distribution in Data Streams with Bounded Deletions [full version]. https://usoop.github.io/file/RUS_full_version.pdf.

[2] Sugam Agarwal, Murali Kodialam, and TV Lakshman. 2013. Traffic engineering in software defined networks. In *Proc. IEEE INFOCOM*. 2211–2219.

[3] Noga Alon, Phillip B Gibbons, Yossi Matias, and Mario Szegedy. 1999. Tracking join and self-join sizes in limited storage. In *Proc. of ACM PODS*. 10–20.

[4] Animashree Anandkumar, Daniel Hsu, and Sham M Kakade. 2012. A method of moments for mixture models and hidden Markov models. In *Proc. of the ACM COLT*, Vol. 23. 33.1–33.34.

[5] Vladimir Braverman, Stephen R Chestnut, Nikita Ivkin, Jelani Nelson, Zhengyu Wang, and David P Woodruff. 2017. Bptree: An l2 heavy hitters algorithm using constant memory. In *Proc. of ACM PODS*. 361–376.
[6] Vladimir Braverman and Rafail Ostrovsky. 2013. Generalizing the layering method of indyk and woodruff: Recursive sketches for frequency-based vectors on streams. In *Proc. of APPROX Workshop*, Vol. 8096. 58–70.
[7] Vladimir Braverman, David Woodruff, and Lin Yang. 2018. Revisiting Frequency Moment Estimation in Random Order Streams. In *Proc. of ICALP*, Vol. 107. 25:1–25:14.
[8] CAIDA. 2016. The CAIDA Anonymized Internet Traces.
[9] Moses Charikar, Kevin Chen, and Martin Farach-Colton. 2004. Finding frequent items in data streams. *Theoretical Computer Science* 312, 1 (2004), 3–15.
[10] Michael P Connolly, Nicholas J Higham, and Theo Mary. 2021. Stochastic rounding and its probabilistic backward error analysis. *SIAM Journal on Scientific Computing* 43, 1 (2021), A566–A585.
[11] Jeffrey Considine, Feifei Li, George Kollios, and John Byers. 2004. Approximate aggregation techniques for sensor databases. In *Proc. of IEEE ICDE*. 449–460.
[12] Graham Cormode and Shan Muthukrishnan. 2005. An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithms* 55, 1 (2005), 58–75.
[13] Antonios Deligiannakis, Nikos Giatrakos, Yannis Kotidis, Vasilis Samoladas, and Alkis Simitsis. 2021. Extreme-Scale Interactive Cross-Platform Streaming Analytics-The INFORE Approach. In *Proc. of SEA-Data@ VLDB Workshop*, Vol. 2929. 7–13.
[14] Cristian Estan and George Varghese. 2002. New directions in traffic measurement and accounting. In *Proc. of ACM SIGCOMM*. 323–336.
[15] Cristian Estan, George Varghese, and Mike Fisk. 2003. Bitmap algorithms for counting active flows on high speed links. In *Proc. of ACM SIGCOMM*. 153–166.
[16] Zhuochen Fan, Ruixin Wang, Yalun Cai, Ruwen Zhang, Tong Yang, Yuhan Wu, Bin Cui, and Steve Uhlig. 2023. OneSketch: A Generic and Accurate Sketch for Data Streams. *IEEE Transactions on Knowledge and Data Engineering* 35, 12 (2023), 12887–12901.
[17] Edward Gan, Jialin Ding, Kai Sheng Tai, Vatsal Sharan, and Peter Bailis. 2018. Moment-Based Quantile Sketches for Efficient High Cardinality Aggregation Queries. In *Proc. of VLDB Endow.*, Vol. 11. 1647–1660.
[18] Minos Garofalakis, Johannes Gehrke, and Rajeev Rastogi. 2016. *Data stream management: processing high-speed data streams*. Springer.
[19] Michael Geller and Pramod Nair. 2018. 5G security innovation with Cisco. *Whitepaper Cisco Public* (2018), 1–29.
[20] Nikos Giatrakos, David Arnu, Theodoros Bitsakis, Antonios Deligiannakis, Minos Garofalakis, Ralf Klinkenberg, Aris Konidaris, Antonis Kontaxakis, Yannis Kotidis, Vasilis Samoladas, et al. 2020. Infore: Interactive cross-platform analytics for everyone. In *Proc. of ACM CIKM*. 3389–3392.
[21] Stefan Heule, Marc Nunkesser, and Alexander Hall. 2013. Hyperloglog in practice: Algorithmic engineering of a state of the art cardinality estimation algorithm. In *Proc. of Springer EDBT*. 683–692.
[22] Rajesh Jayaram and David P Woodruff. 2018. Data streams with bounded deletions. In *Proc. of ACM PODS*. 341–354.
[23] Yingsheng Ji, Zheng Zhang, Xinlei Tang, Jiachen Shen, Xi Zhang, and Guangwen Yang. 2022. Detecting cash-out users via dense subgraphs. In *Proc. of ACM SIGKDD*. 687–697.
[24] Aarati Kakaraparthy, Jignesh M Patel, Brian P Kroth, and Kwanghyun Park. 2022. VIP hashing: adapting to skew in popularity of data on the fly. In *Proc. of VLDB Endow.*, Vol. 15. 1978–1990.
[25] Martin Kiefer, Ilias Poulakis, Eleni Tzirita Zacharatou, and Volker Markl. 2023. Optimistic Data Parallelism for FPGA-Accelerated Sketching. In *Proc. of VLDB Endow.*, Vol. 16. 1113–1125.
[26] Antonios Kontaxakis, Nikos Giatrakos, Dimitris Sacharidis, and Antonios Deligiannakis. 2023. And synopses for all: A synopses data engine for extreme scale analytics-as-a-service. *Information Systems* 116 (2023), 102221.
[27] Abhishek Kumar, Minho Sung, Jun Xu, and Jia Wang. 2004. Data streaming algorithms for efficient and accurate estimation of flow size distribution. *ACM SIGMETRICS Performance Evaluation Review* 32, 1 (2004), 177–188.
[28] Ashwin Lall, Vyas Sekar, Mitsunori Ogihara, Jun Xu, and Hui Zhang. 2006. Data streaming algorithms for estimating entropy of network traffic. *ACM SIGMETRICS Performance Evaluation Review* 34, 1 (2006), 145–156.
[29] Kasper Green Larsen, Rasmus Pagh, and Jakub Tětek. 2021. Countsketches, feature hashing and the median of three. In *International Conference on Machine Learning*. PMLR, 6011–6020.
[30] Alexandru Lavric and Valentin Popa. 2017. Internet of things and LoRa™ low-power wide-area networks: a survey. In *Proc. of IEEE ISSCS*. 1–5.
[31] Jizhou Li, Zikun Li, Yifei Xu, Shiqi Jiang, Tong Yang, Bin Cui, Yafei Dai, and Gong Zhang. 2020. Wavingsketch: An unbiased and generic sketch for finding top-k items in data streams. In *Proc. of ACM SIGKDD*. 1574–1584.
[32] Zaoxing Liu, Antonis Manousis, Gregory Vorsanger, Vyas Sekar, and Vladimir Braverman. 2016. One sketch to rule them all: Rethinking network flow monitoring with univmon. In *Proc. of ACM SIGCOMM*. 101–114.
[33] Zaoxing Liu, Hun Namkung, Georgios Nikolaidis, Jeongkeun Lee, Changhoon Kim, Xin Jin, Vladimir Braverman, Minlan Yu, and Vyas Sekar. 2021. Jaqen: A High-Performance Switch-Native approach for detecting and mitigating volumetric DDoS attacks with programmable switches. In *Proc. of USENIX Security*. 3829–3846.
[34] Nishad Manerikar and Themis Palpanas. 2009. Frequent items in streaming data: An experimental evaluation of the state-of-the-art. *Data & Knowledge Engineering* 68, 4 (2009), 415–430.
[35] Gurmeet Singh Manku and Rajeev Motwani. 2002. Approximate frequency counts over data streams. In *Proc. of VLDB Endow.* 346–357.
[36] Ahmed Metwally, Divyakant Agrawal, and Amr El Abbadi. 2005. Efficient computation of frequent and top-k elements in data streams. In *Proc. of Springer ICDT*, Vol. 3363. 398–412.
[37] Shanmugavelayutham Muthukrishnan et al. 2005. Data streams: Algorithms and applications. *Foundations and Trends® in Theoretical Computer Science* 1, 2 (2005), 117–236.
[38] Pratanu Roy, Arijit Khan, and Gustavo Alonso. 2016. Augmented sketch: Faster and more accurate stream processing. In *Proc. of ACM SIGMOD*. 1449–1463.
[39] Siyuan Sheng, Qun Huang, Sa Wang, and Yungang Bao. 2021. PR-sketch: monitoring per-key aggregation of streaming data with nearly full accuracy. In *Proc. of VLDB Endow.*, Vol. 14. 1783–1796.
[40] James Alexander Shohat and Jacob David Tamarkin. 1950. *The problem of moments*. Vol. 1. American Mathematical Society.
[41] Cha Hwan Song, Pravein Govindan Kannan, Bryan Kian Hsiang Low, and Mun Choon Chan. 2020. FCM-Sketch: Generic Network Measurements with Data Plane Support. In *Proc. of ACM CoNEXT*. 78–92.
[42] Zehua Sun, Huanqi Yang, Kai Liu, Zhimeng Yin, Zhenjiang Li, and Weitao Xu. 2022. Recent advances in LoRa: A comprehensive survey. *ACM Transactions on Sensor Networks* 18, 4 (2022), 1–44.
[43] Lu Tang, Qun Huang, and Patrick PC Lee. 2019. MV-Sketch: A Fast and Compact Invertible Sketch for Heavy Flow Detection in Network Data Streams. In *Proc. of IEEE INFOCOM*. 2026–2034.
[44] Daniel Ting. 2018. Data Sketches for Disaggregated Subset Sum and Frequent Item Estimation. In *Proc. of ACM SIGMOD*. 1129–1140.
[45] Haibo Wang, Chaoyi Ma, Olufemi O Odegbile, Shigang Chen, and Jih-Kwon Peir. 2021. Randomized error removal for online spread estimation in data streaming. In *Proc. of VLDB Endow.*, Vol. 14. 1040–1052.
[46] Haibo Wang, Chaoyi Ma, Olufemi O Odegbile, Shigang Chen, and Jih-Kwon Peir. 2022. Randomized Error Removal for Online Spread Estimation in High-Speed Networks. *IEEE/ACM TON* 31, 2 (2022), 558–573.
[47] Larry Wasserman. 2004. *All of statistics: a concise course in statistical inference*. Vol. 26. Springer.
[48] Kyu-Young Whang, Brad T Vander-Zanden, and Howard M Taylor. 1990. A linear-time probabilistic counting algorithm for database applications. *ACM Transactions on Database Systems* 15, 2 (1990), 208–229.
[49] David P Woodruff and Samson Zhou. 2021. Separations for Estimating Large Frequency Moments on Data Streams. In *Proc. of ICALP*, Vol. 198. 112:1–112:21.
[50] Qingjun Xiao, Xuyuan Cai, Yifei Qin, Zhiying Tang, Shigang Chen, and Yu Liu. 2023. Universal and Accurate Sketch for Estimating Heavy Hitters and Moments in Data Streams. *IEEE/ACM TON* 31, 5 (2023), 1919–1934.
[51] Qingjun Xiao, Yuexiao Cai, Yunpeng Cao, and Shigang Chen. 2023. Accurate and O(1)-Time Query of Per-Flow Cardinality in High-Speed Networks. *IEEE/ACM TON* 31, 6 (2023), 2994–3009.
[52] Qingjun Xiao, Shigang Chen, You Zhou, Min Chen, Junzhou Luo, Tengli Li, and Yibei Ling. 2017. Cardinality estimation for elephant flows: A compact solution based on virtual register sharing. *IEEE/ACM TON* 25, 6 (2017), 3738–3752.
[53] Qingjun Xiao, Yifei Li, and Yeke Wu. 2023. Finding recently persistent flows in high-speed packet streams based on cuckoo filter. *Computer Networks* 237 (2023), 110097.
[54] Qingjun Xiao, Zhiying Tang, and Shigang Chen. 2020. Universal online sketch for tracking heavy hitters and estimating moments of data streams. In *Proc. of IEEE INFOCOM*. 974–983.
[55] Qingjun Xiao, Haotian Wang, and Guannan Pan. 2022. Accurately Identify Time-decaying Heavy Hitters by Decay-aware Cuckoo Filter along Kicking Path. In *Proc. of IEEE/ACM IWQoS*. 1–10.
[56] Qingjun Xiao, You Zhou, and Shigang Chen. 2017. Better with fewer bits: Improving the performance of cardinality estimation of large data streams. In *Proc. of IEEE INFOCOM*. 1–9.
[57] Kaicheng Yang, Sheng Long, Qilong Shi, Yuanpeng Li, Zirui Liu, Yuhan Wu, Tong Yang, and Zhengyi Jia. 2023. SketchINT: Empowering int with towersketch for per-flow per-switch measurement. *IEEE Transactions on Parallel and Distributed Systems* 34, 11 (2023), 2876–2894.
[58] Mingran Yang, Junbo Zhang, Akshay Gadre, Zaoxing Liu, Swarun Kumar, and Vyas Sekar. 2020. Joltik: enabling energy-efficient" future-proof" analytics on low-power wide-area networks. In *Proc. of ACM MobiCom*. 1–14.
[59] Tong Yang, Siang Gao, Zhouyi Sun, Yufei Wang, Yulong Shen, and Xiaoming Li. 2019. Diamond sketch: Accurate per-flow measurement for big streaming data. *IEEE Transactions on Parallel and Distributed Systems* 30, 12 (2019), 2650–2662.

[60] Tong Yang, Junzhi Gong, Haowei Zhang, Lei Zou, Lei Shi, and Xiaoming Li. 2018. Heavyguardian: Separate and guard hot items in data streams. In *Proc. of ACM SIGKDD*. 2584–2593.

[61] Tong Yang, Jie Jiang, Peng Liu, Qun Huang, Junzhi Gong, Yang Zhou, Rui Miao, Xiaoming Li, and Steve Uhlig. 2018. Elastic sketch: Adaptive and fast network-wide measurements. In *Proc. of ACM SIGCOMM*. 561–575.

[62] Tong Yang, Haowei Zhang, Jinyang Li, Junzhi Gong, Steve Uhlig, Shigang Chen, and Xiaoming Li. 2019. HeavyKeeper: an accurate algorithm for finding Top-k elephant flows. *IEEE/ACM TON* 27, 5 (2019), 1845–1858.

[63] Quanwei Zhang, Qingjun Xiao, and Yuexiao Cai. 2023. A generic sketch for estimating super-spreaders and per-flow cardinality distribution in high-speed data streams. *Computer Networks* 237 (2023), 110059.

[64] Fuheng Zhao, Divyakant Agrawal, Amr El Abbadi, and Ahmed Metwally. 2022. SpaceSaving$^{\pm}$: an optimal algorithm for frequency estimation and frequent items in the bounded-deletion model. In *Proc. of VLDB Endow.*, Vol. 15. 1215–1227.

[65] Fuheng Zhao, Punnal Ismail Khan, Divyakant Agrawal, Amr El Abbadi, Arpit Gupta, and Zaoxing Liu. 2023. Panakos: Chasing the Tails for Multidimensional Data Streams. In *Proc. of VLDB Endow.*, Vol. 16. 1291–1304.

[66] Fuheng Zhao, Sujaya Maiyya, Ryan Wiener, Divyakant Agrawal, and Amr El Abbadi. 2021. KLL$^{\pm}$approximate quantile sketches over dynamic datasets. In *Proc. of VLDB Endow.*, Vol. 14. 1215–1227.

[67] Yang Zhou, Tong Yang, Jie Jiang, Bin Cui, Minlan Yu, Xiaoming Li, and Steve Uhlig. 2018. Cold filter: A meta-framework for faster and more accurate stream processing. In *Proc. of ACM SIGMOD*. 741–756.

[68] You Zhou, Yian Zhou, Shigang Chen, and Youlin Zhang. 2018. Highly compact virtual active counters for per-flow traffic measurement. In *Proc. of IEEE INFOCOM*. 1–9.