A APPENDIX

In this section, we show some details which are not provided in the submission due to space limitation, including detailed derivation proofs and reasons of some claims.

A.1 Analysis of the vote replacement

We analyze that the vote replacement [39], which is used in our algorithm framework in the experiment, has a high probability of capturing hot items to show that it is a good eviction algorithm.

Let the events R_i, S_i, and C_i denote that the bucket records the item e_i, the item e_i stays in the bucket, and the bucket captures the item e_i, respectively.

Theorem 4. *For any eviction strategy, the probability of m buckets recording the hot item e_i is*

$$Pr(R_i) = 1 - [1 - Pr(C_i \cdot S_i)]^m \tag{6}$$

Proof. Suppose there is one bucket. When this bucket fails to record item e_i, we get $Pr(fail) = 1 - Pr(C_i \cdot S_i)$. Since each bucket is independent, the probability that all m buckets fail to get item e_i is $Pr(none) = Pr(fail)^m$. Therefore, we get $Pr(R_i) = 1 - Pr(none) = 1 - [1 - Pr(C_i \cdot S_i)]^m$ □

To get $Pr(C_i \cdot S_i)$, we firstly analyze the probability that a bucket captures the items e_i and the items e_i successfully stays in the bucket. Let p_{e_i} denote the probability of item e_i appearing.

Theorem 5. *Assume that there are t items to be processed, f_{e_i} is the frequency of item e_i, and λ is a parameter of the vote replacement. The probability that item e_i stays in the bucket after being captured by the bucket:*

$$\begin{aligned} Pr(S_i|C_i) = \sum_{k=1}^{q} \sum_{l=0}^{t-k} \binom{q}{k} p_{e_i}^k (1 - p_{e_i})^{q-k} * \\ \binom{t-k}{l} (1 - p_{e_i})^l p_{e_i}^{t-k-l} * I(\lambda k > l) \end{aligned} \tag{7}$$

where $q = min(t, f_{e_i})$, and $I(\lambda k > l)$ is an indicator function that is 1 when $\lambda k > l$ and 0 otherwise.

Proof. For the vote replacement strategy, after the bucket captures the item e_i, V_t records the frequency of the item e_i, V_s records the frequency of other items except for item e_i. The item e_i stays in the bucket until $\lambda V_t \le V_s$. Thus, $Pr(S_i|C_i) = Pr(\lambda V_t > V_s) = \sum_{k=1}^{q} \sum_{l=0}^{t-k} Pr(V_t = k) * Pr(V_s = l) * I(\lambda k > l)$, where $q = min(t, f_{e_i})$, $I(\lambda k > l)$ is an indicator function that is 1 when $\lambda k > l$ and 0 otherwise. Since the number of times the item e_i appears follows a binomial distribution $B(t, p_{e_i})$, we get the equation 7. □

Next, we analyze the probability that a bucket captures item e_i.

Theorem 6. *Assume that there are t items to be processed, and N is the number of distinct items. The probability of the bucket capturing item e_i:*

$$Pr(C_i) = p_{e_i} + (1 - p_{e_i})\left(\sum_{j=0, j \ne i}^{N} p_j Pr(e_i \ evicts \ e_j)\right) \tag{8}$$

If the item e_j is in the bucket, after processing t items,

$$\begin{aligned} Pr(e_i \ evicts \ e_j) = p_{e_i} \sum_{k=1}^{q} \sum_{l=0}^{t-k} \binom{q}{k} p_{e_j}^k (1 - p_{e_j})^{q-k} * \\ \binom{t-1-k}{l} (1 - p_{e_j})^l p_{e_j}^{t-1-k-l} * I(\lambda k \le l+1) \end{aligned} \tag{9}$$

where $q = min(t-1, f_{e_j})$, and $I(\lambda k \le l+1)$ is an indicator function that is 1 when $\lambda k \le l+1$ and 0 otherwise.

Proof. There are two situations for capturing item e_i.

(1) The empty bucket captures the item e_i. Since the probability of an item being stored in an empty bucket is independent and affected by its occurrence probability, $Pr(\text{empty bucket captures } e_i) = p_{e_i}$.

(2) Another item occupies the bucket and item e_i evicts that item. If there is the item e_j in the bucket, based on the vote replacement, we get $Pr(\text{item } e_i \text{ evicts item } e_j) = Pr(\lambda V_t \le V_s | \text{the last one is } e_i) * Pr(\text{the last one is } e_i)$. Based on Theorem 5, we can infer and get the equation 9.

Thus, we get $Pr(C_i) = Pr(\text{empty bucket captures } e_i) + Pr(\text{empty bucket fails to capture } e_i) * \sum_{j=0, j \ne i}^{N} Pr(e_j \text{ occupies the bucket}) * Pr(e_i \text{ evicts } e_j) = p_{e_i} + (1 - p_{e_i}) \sum_{j=0, j \ne i}^{N} p_{e_j} Pr(e_i \text{ evicts } e_j)$. □

Based on Theorem 5 and Theorem 6, we can get $Pr(C_i \cdot S_i) = Pr(S_i|C_i) * Pr(C_i)$. Then, according to Theorem 4, we get the probability of capturing hot items using the vote replacement. Since $Pr(R_i)$ increases monotonically as $Pr(C_i \cdot S_i)$ increases, and the higher p_{e_i} is, the higher $Pr(C_i \cdot S_i)$ is, the the vote replacement has a high probability of capturing hot items.

A.2 Setting of T_x

The setting of T_x can be determined theoretically. In the section 5, we have done the experiment on the impact of T_x on accuracy and the results show that setting T_x to a value within the range of 0.5 to 0.7 of w_1 can achieve high accuracy. In other comparative experiments, we set T_x to $0.6321w_1$, which can be obtained theoretically. In this section, we introduce the theory for determining the setting of T_x.

Let event A_i and U denotes the value of the i-th counter in layer L_1 of cold part is 0, and the number of counters whose value is 0.

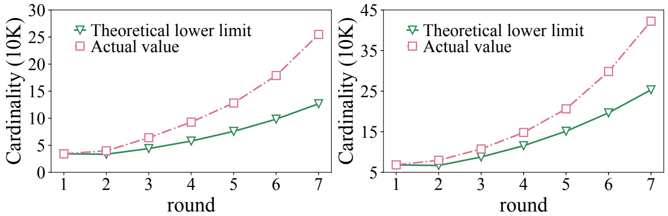
Theorem 7. *Suppose p is the current sampling rate, and n denotes the number of items to be processed. The expected threshold T_x is*

$$E(T_x) = w_1(1 - e^{-\frac{k_1 pn}{w_1}}) \tag{10}$$

where there are k_1 hash functions and w_1 counters in layer L_1.

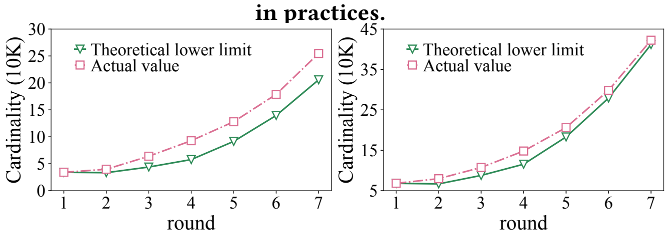
Proof. After n items are processed, $Pr(A_i) = (1 - \frac{1}{w_1})^{k_1 pn}$. Since each counter is independent, $E(U) = \sum_{j=1}^{w_1} Pr(A_j) = w_1(1 - \frac{1}{w_1})^{k_1 pn} = w_1(1 - \frac{1}{w_1})^{w_1 \frac{k_1 pn}{w_1}}$. When both w_1 and n go to infinity, we get $E(U) \approx w_1 e^{-\frac{k_1 pn}{w_1}}$. Thus, $E(T_x) = w_1 - E(U) = w_1(1 - e^{-\frac{k_1 pn}{w_1}})$. □

Since $n \gg w_1$ and $k_1 pn$ is the largest when $p = 1$, based on Theorem 7, we can get $E(T_x) \le w_1(1 - e^{-1}) \le 0.6321w_1$. Therefore, we set T_x to $0.6321w_1$ in our experiments.



(a) w_1 = 102236 (b) w_1 = 204472

Figure 18: Theoretical value of Theorem 1 vs. its real value in practices.



(a) w_1 = 102236 (b) w_1 = 204472

Figure 19: Theoretical value of Theorem 1 vs. its real value in practices, after bridging the gap.

A.3 Experiments of Theorem 1

We have proven the minimum number of distinct items (cardinality) needed in each round in Theorem 1, and We run experiments on Theorem 1 to show its correctness. Figure 18 shows the theoretical lower limit and the actual number of distinct items recorded in each round. The actual value is close to the theoretical value before the 4-th round, but the gap between them increases after that round. This gap is caused by $Pr(A_{i-1}^{=2})$ and $Pr(A_{i-1}^{=3})$ which are ignored when we calculate $Pr(A_i^{=1})$ in Theorem 1. Therefore, we give a complementary equation 11 to make up for the gap, addressing the issue caused by the ignored probabilities.

Let $\Gamma(j, n_i, p_i) = \binom{kn_ip_i}{j}\frac{1}{w_1^j}(1-\frac{1}{w_1})^{kn_ip_i-j}$, we can get $Pr(A_i^{=2,3})$ $= Pr(A_{i-1}^{\leq 1})\sum_{j=2}^{3}\Gamma(j, n_i, p_i) + Pr(A_{i-1}^{=2,3})\sum_{j=1}^{2}\Gamma(j, n_i, p_i) + Pr(A_{i-1}^{=4,5})$ $\sum_{j=0}^{1}\Gamma(j, n_i, p_i) + Pr(A_{i-1}^{=6,7})\Gamma(0, n_i, p_i)$. When n_i and p_i are fixed, $\Gamma(j, n_i, p_i)$ monotonically increases on the interval $j \in [0, \frac{n_i}{2}]$ and monotonically decreases on the interval $j \in [\frac{n_i}{2}, n_i]$. When both w_1 and n_i go to infinity, we get

$$Pr(A_i^{=2,3}) \geq Pr(A_{i-1}^{\leq 1})\sum_{j=2}^{3}\Gamma(j, n_i, p_i) + Pr(A_{i-1}^{=2,3})\sum_{j=1}^{2}\Gamma(j, n_i, p_i) \tag{11}$$

Based on equation 11, we re-run the experiments and figure 19 shows the theoretical number of distinct items recorded in each round, as well as the actual number. After adding the $Pr(A_{i-1}^{=2,3})$, the theoretical value is close to the actual value per round.