Responses to Reviewer ajW8 & Revised Mathematical Analysis

Authors of Paper #190

We sincerely thank the reviewer for carefully reviewing our resubmission and for raising important concerns about the theoretical and empirical aspects of our work. We address each point (W1-W4 & Q1-Q2) in detail below.

W1 and Q1: The Explicit Computation of y

In the current manuscript, we chose not to present the explicit computation of γ (and hence y) in Section 5 due to space constraints and to keep the item of the bound derivations concise. Instead, we introduced γ as the proportion of the N observed items whose frequencies exceed the threshold for e, and then set

$$\lambda = \frac{\gamma N}{2m}$$

accordingly.

In our frequency model, item frequencies follow a power-law distribution with exponent $\alpha > 1$. Let $f_{\min}$ denote the minimum observed frequencies within the window, and f_{e} be the threshold frequency for e. We have the following lemma to estimate γ

Lemma 1 (Proportion of items exceeding threshold in a power-law frequency model)**.** *The proportion γ of items whose frequency exceeds f_{e} in a power-law frequency model is*

$$\gamma = \left(\frac{f_{\mathrm{e}}}{f_{\min}}\right)^{1-\alpha},$$

where $\alpha > 1$, $f_{\min}$ is the minimum observed frequency, and f_{e} is the threshold frequency.

Proof. We assume item frequencies follow a power-law distribution

$$p(f) = Cf^{-\alpha}, \quad f \in [f_{\min}, \infty),$$

where C is the normalization constant:

$$\int_{f_{\min}}^{\infty} Cf^{-\alpha}\, df = 1.$$

For $\alpha > 1$, the integral evaluates to

$$C \cdot \frac{0 - f_{\min}^{1-\alpha}}{1-\alpha} = 1,$$

so

$$C = (\alpha - 1) f_{\min}^{\alpha-1}.$$

The proportion γ is

$$\gamma = \frac{\int_{f_{\mathrm{e}}}^{\infty} Cf^{-\alpha}\, df}{\int_{f_{\min}}^{\infty} Cf^{-\alpha}\, df}.$$

Since the denominator equals 1 by normalization, we have

$$\gamma = \int_{f_{\mathrm{e}}}^{\infty} (\alpha - 1) f_{\min}^{\alpha-1} f^{-\alpha}\, df = (\alpha - 1) f_{\min}^{\alpha-1} \cdot \frac{f_{\mathrm{e}}^{1-\alpha}}{\alpha - 1}.$$

Simplifying gives

$$\gamma = \left(\frac{f_{\mathrm{e}}}{f_{\min}} \right)^{1-\alpha}.$$

□

W2: The Normalization and Parameter Estimates

We thank the reviewer for highlighting this important issue. As explained in our response to W1, we have now provided a normalization for the power-law model that explicitly includes the lower cutoff $f_{\min}$:

$$p(f) = C f^{-\alpha}, \quad C = C = (\alpha - 1) f_{\min}^{\alpha-1}.$$

In that derivation, we have also stated the condition $\alpha > 1$.

To further support our theoretical claims, we will augment the revised manuscript with empirical estimates of α and $f_{\min}$ obtained from our datasets, using maximum-likelihood estimation.

Table 1: CAIDA dataset

Window	$f_{\min}$	α_{MLE}
Window 1	1	3.008939
Window 2	1	2.976420
Window 3	1	2.979181
Window 4	1	2.969790
Window 5	1	2.951388
Window 6	1	2.992581
Window 7	1	2.980728
Window 8	1	2.961826
Window 9	1	2.948427
Window 10	1	2.948217

Table 2: MAWI dataset

Window	$f_{\min}$	α_{MLE}
Window 1	1	6.629547
Window 2	1	6.752943
Window 3	1	6.224208
Window 4	1	6.170716
Window 5	1	6.138736
Window 6	1	6.507459
Window 7	1	6.494920
Window 8	1	6.527997
Window 9	1	6.383061
Window 10	1	6.411964

W3 and Q2: Metric Differences Across Versions

We thank the reviewer for the thoughtful comment. The changes in reported metrics (recall, accuracy, F1) between versions stem solely from the structural optimization of our algorithm. No changes were made to the datasets or parameter settings; all experiments were conducted under identical conditions. The optimization primarily reduced redundant computations and improved memory access patterns, which yielded better wall-clock performance without sacrificing theoretical guarantees.

W4: Independence Assumption and Union Bounds

We now explicitly state the independence assumption when combining Poisson occupancy and power-law order-statistic components:

Proof. Let A be the event that the item e is masked in both candidate buckets after insertion (promotion failure), and B the event that e fails burst detection in the centralized stage (detection failure).

Condition on e being inserted in at least one candidate bucket. Under Assumption3 (independent arrivals) and Assumption1 (ideal hashing), bucket composition in the current window (affecting A) depends only on current arrivals, while burst detection (affecting B) depends on maxima from the previous window. These are based on independent samples over disjoint time intervals, hence A and B are independent.

Therefore:

$$\Pr(A \cap B) = P_{\text{mask}}\, P_{\text{miss,burst}}, \quad \Pr(A \cup B) = P_{\text{mask}} + P_{\text{miss,burst}} - P_{\text{mask}} P_{\text{miss,burst}}.$$

□

1 Mathematical Analysis

(The changes made in the revised Mathematical Analysis according to the comments of Reviewer ajW8 are highlighted with red color.)

1.1 Insertion of Part 1

1.1.1 Time complexity

The insertion algorithm for Part 1 in PBSketch consists of the following steps:

- **Hash Computation:** Given an input item e, two hash functions $h_1(e)$ and $h_2(e)$ are computed to determine the candidate buckets in two arrays $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. Here, calculating two independent hash functions $h_1(e)$ and $h_2(e)$ takes $O(1)$ time.
- **Cell Scan:** For each candidate bucket, n cells are scanned to check whether e already exists, which takes a total of $O(n)$ time.
- **Sorting:** If an update or insertion is triggered, we will use pop sorting to sort the cell, so it will cost $O(n)$.
- **Replacement/Update:** In case of replacement, update the corresponding cell(s) with new values, it costs $O(1)$ time.

Combining these steps, the overall time complexity for inserting an item is dominated by the scanning and sorting process and can be expressed as $O(n)$, where n is the number of cells per bucket. Since n is small and constant in practical deployments, the actual insertion time is $O(1)$ per item.

1.1.2 Error bound

Suppose the following assumptions hold: (1) *Hash Function*: h_1 and h_2 are ideal 32-bit hash functions with uniform distribution; (2) *Frequency Distribution*: item frequencies follow a power-law distribution $P(f) \propto f^{-\alpha}$; (3) *Independence:* arrivals of different items are mutually independent; (4) *Parameter Setting*: m denotes the total number of buckets in each hash table, n the number of cells per bucket, H is the burst frequency threshold, k is the burst ratio threshold, and the target item's frequency satisfies $f^e \geq H$.

A burst item e with frequency f^e may not be recorded in Part 1 due to three main reasons.

First, in the *insertion stage*, a burst item e may fail to be recorded if its candidate buckets are already fully occupied by other items with higher frequencies. To rigorously estimate this probability, we perform the following analysis under the assumption that hash functions are ideal and the distribution of items is sufficiently large and sparse.

Suppose there are N distinct items observed in the window, among which γN items have frequencies higher than the threshold for e. Each such high-frequency item is independently and uniformly mapped into m buckets using 2-hashing. For a target item e, we ask: what is the probability that both of its 2 candidate buckets are fully occupied by high-frequency items and hence e cannot be inserted?

We proceed step by step:

Lemma 1 (Proportion of items exceeding threshold in a power-law frequency model)**.** *The proportion γ of items whose frequency exceeds f_e in a power-law frequency model is*

$$\gamma = \left(\frac{f_e}{f_{\min}}\right)^{1-\alpha},$$

where $\alpha > 1$, $f_{\min}$ is the minimum observed frequency, and f_e is the threshold frequency.

Proof. We assume item frequencies follow a power-law distribution

$$p(f) = Cf^{-\alpha}, \quad f \in [f_{\min}, \infty),$$

where C is the normalization constant:

$$\int_{f_{\min}}^{\infty} Cf^{-\alpha}\, df = 1.$$

For $\alpha > 1$, the integral evaluates to

$$C \cdot \frac{0 - f_{\min}^{1-\alpha}}{1-\alpha} = 1,$$

so

$$C = (\alpha - 1) f_{\min}^{\alpha-1}.$$

The proportion γ is

$$\gamma = \frac{\int_{f_e}^{\infty} Cf^{-\alpha}\, df}{\int_{f_{\min}}^{\infty} Cf^{-\alpha}\, df}.$$

Since the denominator equals 1 by normalization, we have

$$\gamma = \int_{f_e}^{\infty} (\alpha - 1) f_{\min}^{\alpha-1} f^{-\alpha}\, df = (\alpha - 1) f_{\min}^{\alpha-1} \cdot \frac{f_e^{1-\alpha}}{\alpha - 1}.$$

Simplifying gives

$$\gamma = \left(\frac{f_e}{f_{\min}}\right)^{1-\alpha}.$$

□

Lemma 2. *Suppose there are γN items with frequencies higher than a target item e in the observation window, and each item is independently and uniformly hashed into m buckets using 2-hash functions. Let X be the number of such high-frequency items that get hashed into a particular bucket. When N is large and $\gamma N \gg m$, by the law of rare events, X can be approximated by a Poisson distribution with mean*

$$\lambda = \frac{\gamma N}{2m}.$$

This lemma enables us to capture the number of high-frequency items colliding in any given bucket.

Lemma 3. *Given a bucket with n cells, the probability that it receives at least n or more high-frequency items (and thus is fully occupied by them) is*

$$P_{\text{bucket}}(n) = \Pr[X \geq n] = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda}\lambda^k}{k!},$$

where $X \sim \text{Poi}(\lambda)$ and $\lambda = \frac{\gamma N}{2m}$.

Lemma 4. *Let e be an item of interest that has been mapped to 2 certain candidate buckets via its hash functions. If, for all of these 2 buckets, each one happens to be fully occupied by high-frequency items, then e cannot be inserted or recorded. Therefore, the overall probability that e is not recorded (misses all its candidate buckets) can be estimated as*

$$P_{\text{miss}} \approx [P_{\text{bucket}}(n)]^2$$

where $P_{\text{bucket}}(n)$ is defined in Lemma 3.

Second, after an item e is successfully inserted into the candidate bucket (*i.e.*, it is among the recorded items in at least one of its two buckets), it may still fail to be promoted into a hot cell if both of its candidate buckets have at least one item whose frequency is higher than e's current frequency, *i.e.*, e is always out-competed and never occupies a hot position. We now estimate this masking probability.

Lemma 5. *Suppose bucket B has n cells, and γN items in the window have frequencies strictly greater than e. As in Lemma 2, let $X \sim \text{Poi}(\lambda)$ be the (Poisson-approximated) number of such strong items mapped to B, where $\lambda = \frac{\gamma N}{2m}$. The probability that at least one stronger item is present in B is*

$$P_{\text{dom}} = 1 - \Pr[X = 0] = 1 - e^{-\lambda}$$

Lemma 6. *After insertion, an item e will never reside in a hot cell (in either of its two candidate buckets), if and only if, in both buckets, there is at least one item with strictly higher frequency than e. The probability that e is masked by stronger items in both its buckets is*

$$P_{\text{mask}} \approx [P_{\text{dom}}]^2 = \left(1 - e^{-\lambda}\right)^2.$$

Proof. For each candidate bucket, the event that at least one stronger item resides in the bucket is independent (under the hashing and sparsity assumption). e can only be in the hot cell if it is the bucket's top item; otherwise, it's masked. Thus, e can only be "hot" if at least one of its buckets does not have a stronger item.

Therefore, the probability that e is not hot in any bucket equals the probability that both its candidate buckets each have at least one stronger item, hence the formula above. □

Third, we rigorously analyze the probability that e fails to be detected as bursty in the centralized window processing stage, after successful insertion in a hot cell.

Lemma 7. *items were independently inserted into a hot cell, and their frequencies $X_1, X_2, \ldots, X_n$ are independent random variables drawn from a power-law distribution:*

$$F_X(x) = \Pr(X \leq x) = \begin{cases} 0 & x < f_{\min} \\ 1 - \left(\frac{f_{\min}}{x}\right)^{\alpha-1}, & x \geq f_{\min} \end{cases}$$

where $\alpha > 1$, and $f_{\min}$ is the minimum item frequency.

Let

$$F_{pre}^{\max} = \max\{X_1, \ldots, X_n\}$$

be the maximum in the previous window. Then the Cumulative Distribution Function (CDF) of $F_{pre}^{\max}$ is:

$$\Pr\left(F_{pre}^{\max} \le t\right) = [F_X(t)]^n$$

and thus the Probability Density Function (PDF) is:

$$f_{F_{pre}^{\max}}(t) = n\,[F_X(t)]^{n-1} f_X(t)$$

where for $x \ge f_{\min}$,

$$f_X(x) = \frac{dF_X}{dx} = (\alpha - 1) f_{\min}^{\alpha-1} x^{-\alpha}$$

Lemma 8. *Let e be an item with frequency $f \ge H$ in the current window, and k be the burst detection ratio. Then the probability that e's frequency ratio to the previous-window hot cell maximum is less than k, i.e., e fails to be detected as bursty, is:*

$$P_{\text{miss,burst}} = \Pr\left(\frac{f}{F_{pre}^{\max}} < k\right) = \Pr\left(F_{pre}^{\max} > \frac{f}{k}\right) = 1 - \left[F_X\left(\frac{f}{k}\right)\right]^n$$

Proof.

$$\begin{aligned} P_{\text{miss,burst}} &= \Pr\left(f < kF_{pre}^{\max}\right) \\ &= \Pr\left(F_{pre}^{\max} > \frac{f}{k}\right) \\ &= 1 - \Pr\left(F_{pre}^{\max} \le \frac{f}{k}\right) \\ &= 1 - [F_X(f/k)]^n \end{aligned}$$

□

If $f/k \ge f_{\min}$,

$$F_X\left(\frac{f}{k}\right) = 1 - \left(\frac{kf_{\min}}{f}\right)^{\alpha-1}$$

Thus,

$$P_{\text{miss,burst}}(f, n, \alpha, k, f_{\min}) = 1 - \left[1 - \left(\frac{kf_{\min}}{f}\right)^{\alpha-1}\right]^n$$

If $f/k < f_{\min}$, then $F_X(f/k) = 0$, so $P_{\text{miss,burst}} = 1$ (leak is certain).

Lemma 9 (Independence of promotion and detection failures)**.** *Let A be the event that item e suffers* promotion failure *(masked in both candidate buckets after insertion), and let B be the event that e suffers* detection failure *(fails the burst-detection test in the centralized window processing stage).*

Condition on the event that e is successfully inserted in at least one of its candidate buckets. Under Assumption 3 (independent arrivals) and the ideal hashing model (Assumption 1), the bucket composition after insertion depends only on the frequencies of items *assigned to that bucket (affecting A), while the detection phase depends solely on the empirical comparison of e's frequency in the current window to the maxima from the* previous *window (affecting B).*

Since the composition in the current window and the previous-window maxima distribution are generated by independent sampling processes over disjoint time intervals, A and B are independent given successful insertion. Therefore:

$$\Pr(A \cap B) = \Pr(A)\Pr(B) = P_{\text{mask}} \cdot P_{\text{miss,burst}},$$

and the union probability follows:

$$\Pr(A \cup B) = P_{\text{mask}} + P_{\text{miss,burst}} - P_{\text{mask}} P_{\text{miss,burst}}.$$

Overall Error Bound Summary. Combining the above analyses, the probability that a burst item e with frequency $f^e \geq H$ is missed in Part 1 can be comprehensively expressed as follows. The overall miss event occurs if e is not successfully inserted (*insertion failure*), or after successful insertion, it never becomes hot (*promotion failure*), or it is not registered as bursty in the centralized detection phase (*detection failure*). However, as discussed above, the event of insertion failure (Part 1) is strictly stronger than that of promotion failure (Part 2), *i.e.*, an item that cannot be inserted must necessarily fail to become hot.

Furthermore, by Lemma 9, under the assumption that the promotion and detection stages are statistically independent (conditioned on successful insertion), the probability that e is missed after insertion can be given by the union of the two independent events (promotion failure or detection failure):

$$\begin{aligned} P_{\text{miss,total}} &\approx 1 - (1 - P_{\text{mask}})(1 - P_{\text{miss,burst}}) \\ &= P_{\text{mask}} + P_{\text{miss,burst}} - P_{\text{mask}} P_{\text{miss,burst}} \\ &= \left(1 - e^{-\lambda}\right)^2 + \left[1 - \left(1 - \left(\frac{k f_{\min}}{f}\right)^{\alpha-1}\right)^n\right] \\ &\quad - \left(1 - e^{-\lambda}\right)^2 \left[1 - \left(1 - \left(\frac{k f_{\min}}{f}\right)^{\alpha-1}\right)^n\right] \end{aligned}$$

where $\lambda = \frac{\gamma N}{2m}$, P_{mask} is the probability e never becomes hot (Lemma 6), and $P_{\text{miss,burst}}$ is the probability that e is not registered as bursty in the window centralized processing stage (Lemma 8).

1.2 Insertion of Part 2

1.2.1 Time complexity

- **Hash Computation:** For a newly arrived burst item $\langle e, v\rangle$, compute a hash function $h(e)$ to map it to a specific bucket in $\mathcal{D}$. The calculation of $h(e)$ requires $O(1)$ time.
- **Cell Scan:** Scan all n' cells in the target bucket $\mathcal{D}[h(e)]$ to check if an equivalent $\langle e, v\rangle$ already exists, costing $O(n')$ time.

- **Update/Insertion:** If the item exists, increment the r value; if there is an empty cell, insert $\langle e, v, 1 \rangle$ into that cell. Both operations are in $O(1)$ time.
- **Replacement with Probabilistic Eviction:** If the bucket is full, compute the minimum r_{min} among the cells ($O(n')$), decide on replacement with a constant-time probability calculation and, if selected, update the corresponding cell. Overall, this step is dominated by finding r_{min}, which is $O(n')$.

Combining these steps, the total time complexity for inserting a burst item in Part 2 is dominated by the cell scanning and minimum-value search process, expressible as $O(n')$, where n' is the number of cells per bucket in $\mathcal{D}$. Since n' is a small constant in practical settings, the actual insertion time per item is effectively $O(1)$ in real-world deployments.

1.2.2 Error bound

To analyze the error bound in burst tracking for Part 2, let us consider the process where each burst event $\langle e, v \rangle$ is to be inserted into the Part 2 table, which employs m' buckets, each with n' cells, and uses the probabilistic eviction rule detailed previously.

Assume there are M different $\langle e, v \rangle$ pairs, whose burst frequencies follow a power-law distribution f^{-2}.

Guarantee for Frequent Bursty Events Let f_{new} denote the burst count of a newly inserted pair, and f_{min} denote the minimum burst count among the current items in the corresponding bucket. Due to the eviction rule, when $f_{new} \geq 2f_{min}$, the new pair is guaranteed to replace the item with the minimum count after at most $2f_{min}$ failed attempts, because the replacement probability becomes $P = \frac{1}{2f_{min} - C_{fail} + 1}$, and when C_{fail} reaches $2f_{min}$, $P = 1$. Consequently, all bursty events whose burst count is at least twice as large as the current smallest count in their hashed bucket will be eventually inserted and preserved in the Part 2 table.

Guarantee for Top-K Frequent Events Define the burst count of the K-th most frequent pair as f_K. Using the power-law assumption, no more than $2K$ pairs have frequency above $f_K/2$:

$$P(f \leq \frac{f_K}{2}) \approx 1 - \frac{2K}{M}.$$

This implies that, among all pairs, at least a fraction $1 - \frac{2K}{M}$ has a burst count less than $f_K/2$.

Assume that items are assigned to buckets randomly, and let ρ denote the load factor (*i.e.*, probability that a given cell is occupied). For a top-K pair, the insertion succeeds directly if it finds an empty cell $(1-\rho)$, or it collides with a lower-frequency item (probability $\rho(1-\frac{2K}{M})$), in which case probabilistic eviction ensures the higher-frequency item can replace the lower-frequency one, as shown above.

Hence, the lower bound for the insertion probability (for a top-K pair) is:

$$1 - \rho + \rho \left(1 - \frac{2K}{M}\right).$$

Averaging over all possible load factors by integrating ρ from 0 to 1, we get the average lower bound:

$$\int_0^1 \left(1 - \rho + \rho\left(1 - \frac{2K}{M}\right)\right) d\rho = 1 - \frac{K}{M}.$$

That is, at least $1 - \frac{K}{M}$ fraction of top-K bursty events can be made to remain in the Part 2 table under power-law input, thanks to the probabilistic eviction.

Conclusion Therefore, given the probabilistic eviction mechanism, Part 2 provides strong guarantees for retaining all bursty events whose activity is significantly higher than the majority (heavy hitters), and ensures that for top-K events, the expected retention ratio is at least $1 - \frac{K}{M}$.