Appendix

.1 Proof on Lemma 2 and Lemma 3

Lemma 2 $\forall u \in V, p \in H_0, 0 \leq p(u) \leq 1$, have the following relation:

$$\prod_{p \in H_0} (1 - p(u)) \geq 1 - \sum_{p \in H_0} p(u). \tag{1}$$

Proof Let p_i be the distribution of the hash function h_i, then Equation (1) can be expressed as:

$$\prod_{i=0}^{k} (1 - p_i(u)) \geq 1 - \sum_{i=0}^{k} p_i(u). \tag{2}$$

We denote Equation (2) as Ψ. Next we use mathematical induction to prove Ψ, obviously it holds when $k = 0$, we assume that Ψ holds when $k = \alpha - 1$, then we have $\prod\limits_{i=0}^{\alpha-1} (1 - p_i(u)) \geq 1 - \sum\limits_{i=0}^{\alpha-1} p_i(u)$ and we can get:

$$\begin{aligned}
\prod_{i=0}^{\alpha} (1 - p_i(u)) &= (1 - p_\alpha(u)) \prod_{i=0}^{\alpha-1} (1 - p_i(u)) \\
&= \prod_{i=0}^{\alpha-1} (1 - p_i(u)) - p_\alpha(u) \prod_{i=0}^{\alpha-1} (1 - p_i(u)) \\
&\geq 1 - \sum_{i=0}^{\alpha-1} p_i(u) - p_\alpha(u) \prod_{i=0}^{\alpha-1} (1 - p_i(u)) \\
&\geq 1 - \sum_{i=0}^{\alpha} p_i(u).
\end{aligned} \tag{3}$$

Therefore, Ψ holds when $k = \alpha$, this completes the proof. □

Lemma 3 $\forall 0 \leq x \leq 1$, Function $f(x) = \frac{\mathsf{S} \cdot x}{\frac{1}{(1-x)^{\mathsf{S}}} - 1}$ is a convex function.

Proof We rewrite the $f(x)$ as follows:

$$f(x) = \frac{\mathsf{S} \cdot x (1-x)^{\mathsf{S}}}{1 - (1-x)^{\mathsf{S}}} = \frac{\mathsf{S} \cdot (1-x)^{\mathsf{S}}}{\sum\limits_{i=0}^{\mathsf{S}-1} (1-x)^i} \tag{4}$$

Let $\mu = 1 - x$ and $\theta = \mathsf{S}$, so $f(\mu) = \frac{\theta \mu^\theta}{\sum_{i=0}^{\theta-1} \mu^i}$, and we can derive $f'(\mu)$ as follows:

$$f'(\mu) = \theta \frac{\sum\limits_{i=\theta-1}^{2\theta-2} (2\theta - 1 - i) \mu^i}{(\sum\limits_{i=0}^{\theta-1} \mu^i)^2} > 0 \tag{5}$$

Since $f'(x) = \frac{\delta f(\mu)}{\delta \mu} \frac{\delta \mu}{\delta x} = -f'(\mu) < 0$, then we can derive $f''(\mu)$ as follows:

$$\begin{aligned}
f''(\mu) &= \frac{\theta}{(\sum\limits_{i=0}^{\theta-1} \mu^i)^4} \Big((\sum_{i=0}^{\theta-1} \mu^i)^2 \sum_{i=\theta-1}^{2\theta-2} i(2\theta - 1 - i) \mu^{i-1} \\
&\quad - 2 \sum_{i=0}^{\theta-1} \mu^i \sum_{i=1}^{\theta-1} i \mu^{i-1} \sum_{i=\theta-1}^{2\theta-2} (2\theta - 1 - i) \mu^i \Big) \\
&= \frac{\theta}{(\sum\limits_{i=0}^{\theta-1} \mu^i)^3} \Big(\sum_{i=0}^{\theta-1} \mu^i \sum_{i=\theta-1}^{2\theta-2} i(2\theta - 1 - i) \mu^{i-1} \\
&\quad - 2 \sum_{i=0}^{\theta-1} i \mu^i \sum_{i=\theta-1}^{2\theta-2} (2\theta - 1 - i) \mu^{i-1} \Big)
\end{aligned} \tag{6}$$

Next, we compare $\sum\limits_{i=0}^{\theta-1} i\mu^i$ with $\frac{\theta-1}{2} \sum\limits_{i=0}^{\theta-1} \mu^i$,

$$\begin{aligned}
&\frac{\theta - 1}{2} \sum_{i=0}^{\theta-1} \mu^i - \sum_{i=0}^{\theta-1} i \mu^i \\
&= \sum_{i=0}^{\theta-1} (\frac{\theta - 1}{2} - i) \mu^i \\
&= \sum_{i=0}^{\frac{\theta-1}{2}} (\frac{\theta - 1}{2} - i)(\mu^i - \mu^{\theta-1-i})
\end{aligned} \tag{7}$$

Since $0 \leq \mu \leq 1$, we have $\sum\limits_{i=0}^{\theta-1} i\mu^i < \frac{\theta-1}{2} \sum\limits_{i=0}^{\theta-1} \mu^i$. According to Equation (6), we have:

$$\begin{aligned}
f''(\mu) &> \frac{\theta}{(\sum\limits_{i=0}^{\theta-1} \mu^i)^3} \Big(\sum_{i=0}^{\theta-1} \mu^i \sum_{i=\theta-1}^{2\theta-2} i(2\theta - 1 - i) \mu^{i-1} \\
&\quad - (\theta - 1) \sum_{i=0}^{\theta-1} \mu^i \sum_{i=\theta-1}^{2\theta-2} (2\theta - 1 - i) \mu^{i-1} \Big) \\
&= \frac{\theta}{(\sum\limits_{i=0}^{\theta-1} \mu^i)^2} \sum_{i=\theta-1}^{2\theta-2} (i - (\theta - 1))(2\theta - 1 - i) \mu^{i-1}
\end{aligned} \tag{8}$$

Therefore, $f''(\mu) > 0$ and $f''(x) = \frac{\delta^2 f(\mu)}{\delta^2 \mu} (\frac{\delta \mu}{\delta x})^2 + \frac{\delta f(\mu)}{\delta \mu} (\frac{\delta^2 \mu}{\delta^2 x}) = f''(\mu) > 0$. Since $f'(x) < 0, f''(x) > 0$, $f(x)$ is a convex function.

□

.2 Analysis of P'_c

To simplify the analysis, we assume that each bit in Bloom filter is set to 0 with probability p_0 and

1 with probability $1-p_0$. Note that we do not consider the case of *cost exchange*. When all buckets mapped by e_{sk} through all hash functions in H_c are *conflict after adjustment*, we cannot adjust the hash functions of e_{sk}, so we get

$$P'_c = 1 - \prod_{h\in H_c(e_{sk})} (1-(1-p_0^{k-1})^{\chi(h(e_{sk}))}), \quad (9)$$

where $\chi(i)$ represents the number of keys in the i^{th} bucket of Γ. Moreover, according to average value inequality, we have

$$\begin{aligned} 1-P'_c \le& \left(\frac{\mathsf{H}-k-\sum_{h\in H_c}(1-p_0^{k-1})^{\chi(h(e_{sk}))}}{\mathsf{H}-k}\right)^{\mathsf{H}-k} \\ \le& (1-\frac{1}{\mathsf{H}-k}\sum_{h\in H_c}(1-p_0^{k-1})^{\chi(h(e_{sk}))})^{\mathsf{H}-k} \\ \le& (1-\prod_{h\in H_c}(1-p_0^{k-1})^{\frac{\chi(h(e_{sk}))}{\mathsf{H}-k}})^{\mathsf{H}-k}. \quad (10) \end{aligned}$$

It is easy to prove that: $\forall 0<\alpha<1, \beta\in\mathbb{N}, (1-\alpha)^\beta < 1-\alpha^\beta$, which is similar to Lemma 2, then we have

$$\begin{aligned} 1-P'_c <& 1-(1-p_0^{k-1})^{\sum_{h\in H_c}\chi(h(e_{sk}))} \\ P'_c >& (1-p_0^{k-1})^{\sum_{h\in H_c}\chi(h(e_{sk}))}. \quad (11) \end{aligned}$$

Since function $g''(x) = (1-p_0^{k-1})^x$ is a convex function, by the Jensen inequality, we get

$$E(P'_c) > (1-p_0^{k-1})^{E(\sum_{h\in H_c}\chi(h(e_{sk})))}. \quad (12)$$

Let $\psi = \sum_{h\in H_c}\chi(h(e_{sk}))$, and we assume that $\forall h\in H, e_{sk}\in S$, for a certain unit u in V, the probability that u is mapped by e_{sk} through h is only determined by $p(u)$, so we have

$$E(\psi) = E(\sum_{u=1}^{m}\sum_{p\in H_c}\chi(u)p(u)) = E(\sum_{u=1}^{m}\chi(u)\sum_{p\in H_c}p(u)), \quad (13)$$

where $\chi(u) = \mathsf{O}\sum_{p'\in H_0}p'(u)$, for $\forall p_\alpha\in H_0$, $p_\gamma\in H_c$, p_α and p_γ are independent of each other, we have

$$\begin{aligned} E(\psi) =& \sum_{u=1}^{m}\mathsf{O}E(\sum_{p\in H_0}p(u))\cdot E(\sum_{p\in H_c}p(u)) \\ <& \sum_{u=1}^{m}\frac{\mathsf{O}}{4}(\sum_{p\in H}E(p(u)))^2 = \frac{\mathsf{O}\cdot\mathsf{H}^2}{4m}. \quad (14) \end{aligned}$$

Since $0<(1-p_0^{k-1})<1$, then

$$E(P'_c) > (1-p_0^{k-1})^{\frac{\mathsf{O}\cdot\mathsf{H}^2}{4m}}. \quad (15)$$

.3 Proof of Theorem 5

Theorem 5 *If T is the size of CQ and t is the number of Collsion Keys optimized by HABF, we have*

$$E(t) > \frac{T\cdot P'_c(\omega-k^2)}{\omega+T\cdot P'_c\cdot k^2}. \quad (16)$$

Proof We denote HABF$'$ as the HABF that changes operations as follows: no matter whether e_{ck} is optimized successfully or not, we insert a virtual positive key with k randomly selected hash functions into HashExpressor. Let $E'(t)$ be the expected number of collision keys that can be optimized by HABF$'$. It can be seen intuitively that $E(t)\ge E'(t)$.

Next, we analyze $E'(t)$. Let $P^{(i)}$ be the probability that the i^{th} collision key in CQ is optimized by HABF$'$. As per Equation (**??**), we have

$$P^{(i+1)} = P_{ck}(i) \ge P'_c\cdot P_s(i) > P'_c(1-\frac{k(i+1)}{\omega})^k. \quad (17)$$

It is easy to prove that function $g'(i) = (1-\frac{k(i+1)}{\omega})^k$ is a convex function, and P'_c is not related to i as mentioned before. By the Jensen inequality, we have

$$E(P^{(i+1)}) > P'_c\cdot E(g'(i)) > P'_c\cdot g'(E(i)). \quad (18)$$

For HABF$'$, the number of inserted keys in HashExpressor is equal to the number of optimized collision keys, $E(i)=E'(t)$, then we have

$$E(P^{(i+1)}) > P'_c\cdot g'(E'(t)). \quad (19)$$

Lemma 6 For a random variable X_i, $0\le i\le n$, the value of X_i is 0 or 1, the probability expectation of $X_i=1$ is $E(p_i)$, $\forall i,j\in\mathbb{N}, 0\le i,j\le n, i\ne j$, X_i and X_j are independent of each other, we have

$$E(\sum_{i=0}^{n}X_i) = \sum_{i=0}^{n}E(p_i). \quad (20)$$

It is easy to prove Lemma 6 by mathematical induction. As per Equation (17), $P^{(i+1)}$ is only determined by i, so $\forall 0 \leq \alpha, \beta \leq n, \alpha \neq \beta$, $P^{(\alpha)}$ and $P^{(\beta)}$ are independent of each other. By Lemma 6, we get

$$E'(t) = \sum_{i=0}^{T} E(P^{(i)}) > T \cdot P'_c \cdot g'(E'(t)). \tag{21}$$

As per Lemma 2, $g'(E'(t)) = (1 - \frac{k(E'(t)+1)}{\omega})^k \geq 1 - \frac{k^2(E'(t)+1)}{\omega}$, we have $E'(t) > T \cdot P'_c(1 - \frac{k^2(E'(t)+1)}{\omega})$, then

$$E(t) \geq E'(t) > \frac{T \cdot P'_c(\omega - k^2)}{\omega + T \cdot P'_c \cdot k^2}. \tag{22}$$

This completes the proof. □

.4 Analysis of c-HABF performance

In this subsection, we provide the analysis for the expected FPR of c-HABF. Similar to that of HABF, we have

$$E(F^*_{bbf}) = E(F_{bbf}) - \frac{E(t)}{|O|}, \tag{23}$$

where F^*_{bbf} is the FPR of Blocked Bloom filter after optimization and $|O|$ is the number of negative keys. Let t_i be the number of optimized collision keys of i^{th} block, then we have

$$E(F^*_{bbf}) = E(F_{bbf}) - \frac{\sum_{i=0}^{l} E(t_i)}{|O|}, \tag{24}$$

We denote T as the size of CQ and T_i as the number of collision keys for different blocks. The calculation of $E(t_i)$ in Equation (24) is difficult since the number of collision keys fluctuates from different blocks, and the probability of different blocks that the adjusted hash functions can be inserted into the Blocked HashExpressor is not independent. We first consider a worse design where Each block in Blocked HashExpressor also matches one block in Blocked Bloom filter without sharing space, then the cell number of each HashExpressor block is $\frac{\omega}{l}$. According to Theorem 5, we will have

$$E(t_i) = f(P'_c, T_i) = \frac{T_i \cdot P'_c(\frac{\omega}{l} - k^2)}{\frac{\omega}{l} + T_i \cdot P'_c \cdot k^2}. \tag{25}$$

Here we use approximate estimation, *i.e.*, suppose that each of the T collision keys comes from any of the blocks with equal probabilities. Let X' be the random variable for the number of collision keys mapped into a particular block, then X' also follows the binomial distribution, $Bino(T, \frac{1}{l})$. Hence, we have

$$E(t_i) = \sum_{x=0}^{T} (\binom{n}{i} (\frac{1}{l})^x (1 - \frac{1}{l})^{T-x}) \cdot f(P'_c, x). \tag{26}$$

The distribution of X' can be approximated by the Poisson distribution with the parameter $\frac{T}{l}$ when T is large, namely

$$E(t_i) = \sum_{x=0}^{\infty} Poisson(x, \frac{T}{l}) \cdot f(P'_c, x). \tag{27}$$

The design without sharing HashExpressor blocks will have fewer optimized collision keys for the higher insertion failure probability. Therefore, base on Equation (24) and Equation (27), we have

$$E(F^*_{bbf}) < E(F_{bbf}) - \frac{l}{|O|} \sum_{x=0}^{\infty} Poisson(x, \frac{T}{l}) \cdot f(P'_c, x). \tag{28}$$