

Supplementary Materials of CuckooDuo

1 Pseudocode

Algorithm 1: Insertion workflow of CuckooDuo

Input: An item (key-value pair) $e = \langle key, value \rangle$

```

1 if  $\mathcal{I}_1[h_1(key)]$  or  $\mathcal{I}_2[h_2(key)]$  has empty slot then
2   Insert  $FP(key)$  into the empty slot in CuckooIndex;
3   Insert  $e$  into the corresponding slot in CuckooVault;
4   return;
5  $Q :=$  An empty queue used for BFS;
6 for each fingerprint  $f$  in  $\mathcal{I}_1[h_1(key)]$  and  $\mathcal{I}_2[h_2(key)]$  do
7   Push  $\langle f, path = f \rangle$  into the tail of  $Q$ ;
8 while  $Q.not\_empty$  do
9   Pop  $\langle f, path \rangle$  from the front of  $Q$ ;
10  if length of path exceeds  $L$  then
11    return failure;
12  if  $f$  is stored in  $\mathcal{I}_1[h_1(f)]$  then
13     $h_2(f) = (h_1(f) + hash(f)) \% m$ ;
14    if  $\mathcal{I}_2[h_2(f)]$  has empty slot then
15      Read all items on path from CuckooVault;
16      Write items on path (including  $e$ ) into CuckooVault;
17      Write FPs on path (including  $FP(key)$ ) into CuckooIndex;
18      return success;
19    else if  $\mathcal{I}_2[h_2(f)]$  is full then
20      for each fingerprint  $f'$  in  $\mathcal{I}_2[h_2(f)]$  do
21         $path' \leftarrow (path, f')$ ;
22        Push  $\langle f', path' \rangle$  into the tail of  $Q$ ;
23  else if  $f$  is stored in  $\mathcal{I}_2[h_2(f)]$  then
24     $h_1(f) = (h_2(f) - hash(f)) \% m$ ;
25    Check  $\mathcal{I}_1[h_1(f)]$  and perform similar operations as above;
26 return failure;

```

2 Proof for Theorem IV.1

Theorem IV.1. Consider a basic CuckooDuo under the load factor of α . Let X_α be the number of items failed to be inserted into CuckooDuo due to fingerprint collisions. We have

$$\mathbb{E}(X_\alpha) \approx 2md^2\alpha^2/2^f \leq 4md^2\alpha^2/2^f = O\left(\frac{md^2\alpha^2}{2^f}\right)$$

where m is the number of buckets in each bucket array, d is the bucket size, and f is the length of fingerprints (in bits).

Proof. Consider an incoming item e . The probability that its fingerprint collides with the fingerprint of another item is $1/2^f$. For a CuckooDuo under the load factor of α' , the expected number of items in the two candidate buckets of e is $2d\alpha'$. Therefore, the expected probability of item e experiencing a fingerprint collision is

$$\mathcal{P} = 1 - (1 - 1/2^f)^{2d\alpha'} \approx 2d\alpha'/2^f$$

For the CuckooDuo under the load factor of α , we can calculate the expectation of X_α by integrating \mathcal{P} over the number of inserted items x as

$$\mathbb{E}(X_\alpha) = \int_0^{2dm\alpha} \alpha' \cdot \frac{2d}{2^f} dx = \int_0^{2dm\alpha} \frac{x}{2dm} \cdot \frac{2d}{2^f} dx = \frac{2md^2\alpha^2}{2^f}$$

By fixing the load factor $\alpha' = \alpha$ during integration, we can also derive an upper bound as

$$\mathbb{E}(X_\alpha) \leq 4md^2\alpha^2/2^f$$

□

3 Proof for Theorem IV.2

Theorem IV.2. Consider a CuckooDuo with Dual-Fingerprint optimization. Let X be the number of items failed to be inserted into CuckooDuo due to fingerprint collisions. We have

$$\mathbb{E}(X) \leq \frac{4md(d+1)(d-1)}{3 \cdot 2^{2f}} = O\left(\frac{md^3}{2^{2f}}\right)$$

where m is the number of buckets in each bucket array, d is the bucket size, and f is the length of fingerprints (in bits).

Proof. Let d_1 and d_2 be number of slots in each bucket of \mathcal{I}_1 using $FP_1(\cdot)$ and $FP_2(\cdot)$ respectively ($d = d_1 + d_2$). We assume $d_1 \geq 2$ and $d_2 \geq 2$.

Consider a certain bucket $\mathcal{I}_1[i]$ in the first bucket array. If the items in $\mathcal{I}_1[i]$ have only one fingerprint collision (either in FP_1 or FP_2), the collision can definitely be resolved through our *Dual-Fingerprint* adjustment. In the following, we assume that fingerprint collision occurs uniformly across all buckets. This assumption will lead to an upper bound of the failure probability because our *Dual-Fingerprint* algorithm can effectively resolve many collisions.

Let \mathcal{P} be the probability that there exists fingerprint collisions in $\mathcal{I}_1[i]$ that cannot be resolved through *Dual-Fingerprint* adjustment. The upper bound of \mathcal{P} can be written as $\overline{\mathcal{P}} = 1 - \mathcal{P}_0 - \mathcal{P}_1$, where \mathcal{P}_j is the probability that j fingerprint collisions happen in the certain bucket $\mathcal{I}_1[i]$ ($j = 0$ means no fingerprint collision happens).

We derive \mathcal{P}_0 and \mathcal{P}_1 as

$$\mathcal{P}_0 = \left(\prod_{j=0}^{d-1} \left(1 - \frac{j}{2^f} \right) \right)^2$$

$$\mathcal{P}_1 = 2 \left(\prod_{j=0}^{d-1} \left(1 - \frac{j}{2^f} \right) \right) \cdot \binom{d}{2} \cdot \frac{1}{2^f} \cdot \left(\prod_{j=0}^{d-2} \left(1 - \frac{j}{2^f} \right) \right)$$

where $\binom{d}{2} = \frac{d!}{2!(d-2)!}$ is the combination number.

Without loss of generality, we assume that $d < d^2 < d^3 \ll 2^f$. By using mathematical analysis techniques, we can get:

$$\begin{aligned} \mathcal{P}_0 &= 1 - \frac{d(d-1)}{2^f} + \frac{d^2(d-1)^2 - 2/3 \cdot d(d-1)(2d-1)}{2^{2f}} + o\left(\frac{d^3}{2^{2f}}\right) \\ \mathcal{P}_1 &= \frac{d(d-1)}{2^f} - \frac{d(d-1)^3}{2^{2f}} + o\left(\frac{d^3}{2^{2f}}\right) \\ \overline{\mathcal{P}} &= 1 - \mathcal{P}_0 - \mathcal{P}_1 = \frac{d(d-1)(d+1)}{3 \cdot 2^{2f}} + o\left(\frac{d^3}{2^{2f}}\right) \end{aligned}$$

Recall that we assume $d^3 \ll 2^f$, meaning that $\overline{\mathcal{P}}$ is very small. Therefore, for each item in bucket $\mathcal{I}_1[i]$, the probability that it has a fingerprint collision with another item in $\mathcal{I}_1[i]$ and the collision cannot be resolved with *Dual-Fingerprint* adjustment is \mathcal{P}/d .

As each item might collide with all items in its two candidate bucket, the upper bound of the probability that an item experiencing an unresolvable fingerprint collision is $2\overline{\mathcal{P}}/d$. Finally, as there are at most $2dm$ items in CuckooDuo, the expectation of the number of items with unresolvable collisions satisfies that

$$\mathbb{E}(X) \leq 2dm \cdot \frac{2\overline{\mathcal{P}}}{d} \leq 2dm \cdot \frac{2(d+1)(d-1)}{3 \cdot 2^{2f}} = \frac{4md(d+1)(d-1)}{3 \cdot 2^{2f}}$$

□

4 Proof for Theorem IV.3

Theorem IV.3. Consider a basic CuckooDuo under the load factor of α . Let Y_α be the number of items failed to be inserted into CuckooDuo due to BFS failure (i.e., the length of kick-out path exceeds the predefined threshold L). We have

$$\mathbb{E}(Y_\alpha) \approx 2md \int_0^\alpha \frac{\beta(r)^{2\Sigma_{i=0}^L d^i}}{1 - \beta(r)^{2\Sigma_{i=0}^L d^i}} dr$$

where m is the number of buckets in each bucket array, d is the bucket size, L is the maximum length of kick-out path, and $\beta(r)$ is the ratio of full buckets under the load factor of r .

Proof. Consider a CuckooDuo under the load factor of r . We define its full bucket ratio as $\beta(r) := \frac{t(r)}{2m}$, where $t(r)$ is the number of full buckets in CuckooDuo. Here, $\beta(r)$ and $t(r)$ are two functions of load factor r .

For an incoming new item e , the probability that it cannot be directly inserted into one of its two candidate buckets is $\mathcal{P}_0 = \beta(r)^2$.

During a BFS process with the maximum kick-out path length of L , there are $2\Sigma_{i=0}^L d^i$ buckets to be checked. We assume the BFS process randomly visits each bucket. The probability that item e cannot be inserted into CuckooDuo through the BFS process is

$$\mathcal{P}_L = \beta(r)^{2\Sigma_{i=0}^L d^i}$$

which is also the probability that the BFS cannot find a non-full bucket.

For the CuckooDuo under the load factor of α , we can calculate the expectation of Y_α by integrating over load factor r .

$$\mathbb{E}(Y_\alpha) \approx 2md \int_0^\alpha \frac{\mathcal{P}_L}{1 - \mathcal{P}_L} dr = 2md \int_0^\alpha \frac{\beta(r)^{2\Sigma_{i=0}^L d^i}}{1 - \beta(r)^{2\Sigma_{i=0}^L d^i}} dr$$

□