On the insertion time of random walk cuckoo hashing

Alan Frieze∗and Tony Johansson†
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA15213
U.S.A.

Abstract

We show that if the number of hash functions $d = O(1)$ is sufficiently large, then the expected insertion time of Random Walk Cuckoo Hashing is $O(1)$ per item.

1 Introduction

Our motivation for this paper comes from Cuckoo Hashing (Pagh and Rodler [9]). Briefly each one of $n$ items $x \in L$ has $d$ possible locations $h_1(x), h_2(x), \ldots, h_d(x) \in R$, where $d$ is typically a small constant and the $h_i$ are hash functions, typically assumed to behave as independent fully random hash functions. (See [8] for some justification of this assumption.)

We assume each location can hold only one item. Items are inserted consecutively and when an item $x$ is inserted into the table, it can be placed immediately if one of its $d$ locations is currently empty. If not, one of the items in its $d$ locations must be displaced and moved to another of its $d$ choices to make room for $x$. This item in turn may need to displace another item out of one of its $d$ locations. Inserting an item may require a sequence of moves, each maintaining the invariant that each item remains in one of its $d$ potential locations, until no further evictions are needed.

We now give the formal description of the mathematical model that we use. We are given two disjoint sets $L = \{v_1, v_2, \ldots, v_n\}$, $R = \{w_1, w_2, \ldots, w_m\}$. Each $v \in L$ independently chooses a set $N(v)$ of $d \geq 2$ random neighbors in $R$. We assume for simplicity that this selection is done with replacement. This provides us with the bipartite cuckoo graph $\Gamma$. Cuckoo Hashing can be thought of as a simple algorithm for finding a matching $M$ of $L$ into $R$ in $\Gamma$. In the context of hashing, if $\{x, y\}$ is an edge of $M$ then $y \in R$ is a hash value of $x \in L$.

Cuckoo Hashing constructs $M$ by defining a sequence of matchings $M_1, M_2, \ldots, M_n$, where $M_k$ is a matching of $L_k = \{v_1, v_2, \ldots, v_k\}$ into $R$. We let $R_k$ denote the vertices of $R$ that are covered by $M_k$ and define the function $\phi_k : L_k \to R_k$ by asserting that $M_k = \{\{v, \phi_k(v)\} : v \in L_k\}$. We obtain $M_k$ from $M_{k-1}$ by finding an augmenting path $P_k$ in $\Gamma$ from $v_k$ to a vertex in $R_{k-1} = R \setminus R_{k-1}$.

∗Research supported in part by NSF Grants DMS1362785, CCF1522984
†Research supported in part by NSF Grant DMS1362785
This augmenting path $P_k$ is obtained by a random walk. To begin we obtain $M_1$ by letting $\phi_1(v_1)$ be a random member of $N(v_1)$, the neighbors of $v_1$. Having defined $M_k$ we proceed as follows: Steps 1 – 4 constitute round $k$.

**Algorithm** insert:

**Step 1** $x \leftarrow v_k; M \leftarrow M_{k-1}$;

**Step 2** If $S_k(x) = N(x) \cap \bar{R}_{k-1} \neq \emptyset$ then choose $y$ randomly from $S_k(x)$ and let $M_k = M \cup \{\{x,y\}\}$, else

**Step 3** Choose $y$ randomly from $N(x)$;

**Step 4** $M \leftarrow M \cup \{\{x,y\}\} \setminus \{y,\phi_{k-1}(y)\}; x \leftarrow \phi_{k-1}(y)$; goto Step 2.

Our interest here is in the expected time for insert to complete a round. For large $d$, we will improve on the results of Frieze, Melsted and Mitzenmacher [5], Fountoulakis, Panagiotou and Steger [2], Fotakis, Pagh, Sanders and Spirakis [3]. Mitzenmacher [7] gives a survey on Cuckoo Hashing and Problem 1 of the survey asks for the expected insertion time.

Frieze and Melsted [4], Fountoulakis and Panagiotou [1] give information on the relative sizes of $L, R$ needed for there to exist a matching of $L$ into $R$ w.h.p.

We will prove the following theorem:

**Theorem 1** Suppose that $n = (1 - \varepsilon)m$ where $\varepsilon$ is a fixed positive constant. Then there exists a constant $d_\varepsilon$ such that if $d \geq d_\varepsilon$ then w.h.p. $\Gamma$ is such that over the random choices in Steps 2,3,

$$E(|P_k|) \leq 2$$

for $k = 1, 2, \ldots, n$. (1)

Here $|P_k|$ is the length (number of edges) of $P_k$.

# 2 Proof of Theorem 1

We will w.l.o.g. only prove (1) for the case $k = n$. This is valid since we will see that our claims hold a fortiori for smaller $k$.

We first observe that if $R_{k-1} = \{y_1, y_2, \ldots, y_{k-1}\}$ then

$$y_k \text{ is chosen uniformly from } \bar{R}_{k-1}$$

and is independent of the graph $\Gamma_{k-1}$ induced by $L_{k-1} \cup R_{k-1}$. This is because we can expose $\Gamma$ along with the algorithm. When we start the construction of $M_k$ we expose the neighbors of $v_k$ one by one. In this way we either determine that $S_k(v_k) = \emptyset$ or we expose a random member of $S_k(v_k)$ without revealing any more of $N(v_k)$. In general, in Step 2, we have either exposed all the neighbors of $x$ and these will necessarily be in $R_{k-1}$. Or, we can proceed to expose the unexposed neighbors of $x$ until either (i) we determine that $S_k(x) = \emptyset$ and we choose a random member of $N(x)$ or (ii) we find a neighbor of $x$ that is a random member of $\bar{R}_{k-1}$. Thus $R_{n-1}$ is a random subset of $R$. 2
Let 
\[ B = \{ v \in L : N(v) \cap R_{n-1} = \emptyset \}. \]

If \( x \not\in B \) in Step 2 of INSERT then we will have found \( P_n \).

Let \( P = (x_1, y_1, x_2, y_2, \ldots, x_t) \) be a path in \( \Gamma \), where \( x_1, x_2, \ldots, x_t \in L \) and \( y_1, y_2, \ldots, y_{t-1} \in R \). We say that \( P \) is interesting if \( x_1, x_2, \ldots, x_t \in B \). We note that if the path \( P_n = (x_0 = v_n, y_0, x_1, y_1, \ldots, x_{t-1}, y_{t-1}, y_{t+1}) \) then \( Q_n = (x_1, y_1, x_2, y_2, \ldots, x_t) \) is interesting. Indeed, we must have \( x_i \in B, 1 \leq i \leq t \), else INSERT would have chosen \( y_i \in R_{n-1} \subseteq R_{x_i} \) and completed the round.

Let \( \nu_t \) denote the number of interesting paths with \( 2t - 1 \) vertices.

**Lemma 2** Given \( A_0 \) and \( d \) sufficiently large,
\[ \mathbb{P}(2 \leq \ell \leq A_0 \log \log n : \nu_\ell \geq n) = o(n^{-2}). \] (3)

The bound \( o(n^{-2}) \) is sufficient to deal with the insertion of \( n \) items.

Before proving the lemma, we show how it can be used to prove Theorem 1. We will need the following claims:

**Claim 3** Let \( \Delta \) denote the maximum degree in \( \Gamma \). Then for any \( t \geq \log n \) we have \( \mathbb{P}(\Delta \geq t) \leq e^{-t} \).

**Proof of Claim:** If \( v \in L \) then its degree \( \text{deg}(v) = d \). Now consider \( w \in R \). Then for \( t \geq \log n \),
\[ \mathbb{P}(\exists w \in R : \text{deg}(w) \geq t) \leq m \left( \frac{dn}{t} \right) \frac{1}{m^t} \leq m \left( \frac{de}{t} \right)^t \leq e^{-t}. \]

**End of proof of Claim**

We will make use of the following simple modification of the Azuma-Hoeffding inequality.

**Lemma 4** Let \( Z = Z(X_1, X_2, \ldots, X_M) \geq 0 \) where \( X_1, X_2, \ldots, X_M \) are independent random variables. Let \( E = E(X_1, X_2, \ldots, X_M) \) be an event. Suppose that \( Z \leq A_0 \) always and that if \( E \) occurs, then changing a single \( X_i \) can only change \( Z \) by at most \( A_1 \). Then, for any \( t > A_0 \mathbb{P}(E) \) we have
\[ \mathbb{P}(Z \geq E(Z) + t) \leq \exp \left\{ - \frac{(t - A_0 \mathbb{P}(E))^2}{M A_1^2} \right\} + \mathbb{P}(E). \]

**Proof** We have
\[ \mathbb{P}(Z \geq E(Z) + t) = \mathbb{P}(Z1_E \geq E(Z) + t) + \mathbb{P}(Z1_{\bar{E}} \geq E(Z) + t) \leq \mathbb{P}(Z1_E \geq E(Z1_E) + u) + \mathbb{P}(\bar{E}) \] (4)

where
\[ u = E(Z) - E(Z1_E) + t \geq E(Z1_{\bar{E}})(\mathbb{P}(E) - 1) + t \geq t - A_0 \mathbb{P}(E). \] (5)

Applying the Azuma-Hoeffding inequality (more precisely, the special case referred to as McDiarmid’s inequality) we get
\[ \mathbb{P}(Z1_E \geq E(Z1_E) + u) \leq \exp \left\{ - \frac{u^2}{M A_1^2} \right\}. \] (6)

The lemma follows after using (5) and (6) in (4).
Claim 5  With probability $1 - o(n^{-2})$, $\Gamma$ contains at most $n^{1/2+o(1)}$ cycles of length at most $\Lambda = (\log \log n)^2$.

Proof of Claim: The expected number of cycles of length at most $2\ell = \Lambda$ is bounded by

$$\sum_{k=2}^{\ell} \binom{n}{k} \binom{m}{k} (k!)^2 \left( \frac{d}{m} \right)^{2k} \leq \sum_{k=2}^{\ell} d^{2k} = n^{o(1)}.$$ 

Let $C$ denote the number of cycles of length at most $\Lambda$. We apply Lemma 4 to $C$ with $\mathcal{E} = \{\Delta \leq \lambda = (\log n)^2\}$, $A_0 = m^\Lambda$ and $A_1 = \lambda^\Lambda$ and $t = n^{1/2}A_0 \log n = n^{1/2+o(1)}$. We use Claim 3 to bound $P(\mathcal{E})$.

End of proof of Claim

These two claims imply the following:

With probability $1 - o(n^{-2})$ there are at most $n^{1/2+o(1)}(3 \log n)^{2\ell} = n^{1/2+o(1)}$ vertices within distance $t$ most $2A_0 \log \log n$ of a cycle of length at most $\Lambda = (\log \log n)^2$. (7)

Continuing with the proof of Theorem 1 we will need the following result from [5]: We phrase Claim 10 of that paper in our current terminology.

Claim 6  There exists a constant $a > 0$ such that for any $v \in L_{n-1}$, the expected time for INSERT to reach $R_{n-1}$ is $O((\log n)^a)$.

Now let $p_{\ell}$ denote the probability that INSERT requires at least $\ell$ rounds to insert $v_n$. We prove the theorem by showing that

$$E(|P_n|) = 1 + 2 \sum_{\ell=2}^{\infty} p_{\ell} \leq 2.$$  

(We only need to verify the inequality.)

We observe that if $v_{n-1}$ has no neighbor in $R_{n-1}$ and has no neighbor in a cycle of length at most $\Lambda$ then for some $\ell \leq A_0 \log \log n$, the first $2\ell-1$ vertices of $P_n$ follow an interesting path. Hence,

$$\sum_{\ell=2}^{A_0 \log \log n} p_{\ell} \leq O(n^{-1/2+o(1)}) + \sum_{\ell=2}^{A_0 \log \log n} \frac{\nu_{\ell}}{n(d-1)^{\ell}} \leq O(n^{-1/2+o(1)}) + \sum_{\ell=2}^{A_0 \log \log n} \frac{n}{n(d-1)^{\ell}} \leq \frac{1}{3}.$$  

(9)

Explanation of (9): Following (7), we find that the probability $v_n$ is within $2A_0 \log \log n$ of a cycle of length at most $\Lambda$ is bounded by $n^{-1/2+o(1)}$. The $O(n^{-1/2+o(1)})$ term accounts for $v_n$ choosing a vertex close to a short cycle and for the probability of there being many interesting paths, see (3). Failing this, we have divided the number of interesting paths of length $2\ell-1$ by the number of equally likely walks $n(d-1)^{\ell}$ that INSERT could take. To obtain $n(d-1)^{\ell}$ we argue as follows. We carry out the following thought experiment. We run our walk for $\ell$ rounds regardless. If we manage to choose $y \in R_{n-1}$ then instead of stopping, we move to $y$ and continue. In this way we will in fact be $n(d-1)^{\ell}$ equally likely walks. In our thought experiment we choose one of these walks at random, whereas in the execution of the algorithm we only proceed as far the first time we reach $R_{n-1}$. Finally, for the algorithm to take at least $\ell$ rounds, it must choose an interesting path of length at least $2\ell-1$. 


Note next that
\[ P_{A_0 \log \log n} \leq O(n^{-1/2+o(1)}) + 3^{-A_0 \log \log n}. \]

It follows that
\[ \sum_{\ell = A_0 \log \log n}^{(\log n)^{A_0}} p_{\ell} \leq \sum_{\ell = A_0 \log \log n}^{(\log n)^{A_0}} p_{A_0 \log \log n} \leq \frac{1}{3}. \tag{10} \]

It follows from Claim 6 that for any integer \( \rho \geq 1 \),
\[ \mathbb{P}(|P_n| \geq \rho (\log n)^{2a}) \leq \frac{1}{(\log n)^{\rho a}}. \tag{11} \]

Indeed, we just apply the Markov inequality every \( (\log n)^{2a} \) steps to bound \( |P_n| \) by a geometric random variable.

It follows from (11) that
\[ \sum_{\ell \geq 3(\log n)^{2a}} p_{\ell} \leq \sum_{\ell = 3(\log n)^{2a}}^{\infty} p_{\ell} \leq \sum_{\rho = 3}^{\infty} \frac{1}{(\log n)^{\rho a - 2a}} = o(1). \tag{12} \]

Theorem 1 now follows from (8), (9), (10) and (12), if we take \( A_0 > 2a \).

### 2.1 Proof of Lemma 2

We apply Lemma 4 to \( \nu_{\ell} \) to argue that
\[ \mathbb{P}(\nu_{\ell} \geq \mathbb{E}(\nu_{\ell}) + n^{3/4}) \leq 2e^{-(\log n)^2}. \tag{13} \]

We let \( \mathcal{E} = \{ \Delta \leq \lambda = (\log n)^2 \} \) as before, \( A_0 = m^{2\ell} \), \( A_1 = \lambda^{2\ell} \) and \( t = n^{3/4} \). We use Claim 3 to bound \( \mathbb{P}(\mathcal{E}) \). The bound on \( A_1 \) follows from the fact that an edge can be in at most \( \Delta^{2\ell} \) interesting paths.

It follows from (13) that to finish the proof, all we need to show is that
\[ \mathbb{E}(\nu_{\ell}) \leq \frac{n}{2}. \tag{14} \]

**Claim 7** Let
\[ \mathcal{B} = \left\{ |B| \geq ne^{-\varepsilon d/2} \right\}. \]

Then
\[ \mathbb{P}(\mathcal{B}) = O(\varepsilon^{-\Omega(n^{1/4})}). \tag{15} \]

**Proof of Claim:**

We are aiming for a small probability in (15) and we need to deal with some vertices in \( L \) choosing the same vertex twice. So let \( B_0 = \{ v \in L : |N(v)| < d \} \). Then we have
\[ \mathbb{P}(|B_0| \geq n^{1/4}) \leq \mathbb{P} \left( \text{Bin} \left( n, \frac{d}{m} \right) \geq n^{1/4} \right) \leq e^{-n^{1/4}}. \]
Next let
\[ K = \{ k : \text{ round } k \text{ does not end immediately in Step 2 with } x = v_k \}. \]
Then, \( \mathbb{P}(k \in K) \leq \left( \frac{k}{m} \right)^d \) and this holds for each value of \( k \) independently and so
\[
\mathbb{E}(|K|) \leq \sum_{k=1}^{n} \left( \frac{k}{m} \right)^d \leq n(1 - \varepsilon)^d \leq ne^{-\varepsilon d}.
\]

Now \( |K| \) is the sum of independent \( \{0, 1\} \) random variables and so Hoeffding’s theorem [6] implies that
\[
\mathbb{P}(|K| \geq 2ne^{-\varepsilon d}) = O(e^{-\varepsilon_1 n}) \text{ for some constant } \varepsilon_1 > 0.
\]

Now if \( B_1 = \{ k \in B : \exists \ell \neq k \text{ s.t. round } \ell \text{ ends with } x = v_k \} \) then \( |B_1| \leq K \). But if \( B_2 = B \setminus (K \cup B_0 \cup B_1) \) then \( k \in B_2 \) only if the random choices \( y_1, y_2, \ldots, y_{n-1} \) include all the neighbors of \( v_k \). One of these choices \( w_0 \) say, will be made by \( v_k \) itself in round \( k \). It follows from (2) that, if \( w_1, w_2, \ldots, w_{d-1} \) are the other choices then
\[
\mathbb{P}(v_k \in B_2) = d \prod_{i=1}^{d-1} \mathbb{P}(w_i \in R_{n-1} \mid w_1, w_2, \ldots, w_{i-1} \in R_{n-1}) \leq d \left( 1 - \left( \frac{1}{m} \right) \right)^{d-1} \leq 2d(1 - \varepsilon)^{d-1},
\]
since \( (1 - \frac{1}{m})^n \geq \frac{n}{m} \). The factor \( d \) comes from the \( d \) choices of \( w_0 \).

**Explanation:** \( \mathbb{P}(w_i \notin R_{n-1} \mid w_1, w_2, \ldots, w_{i-1} \in R_{n-1}) \geq (1 - \frac{1}{m})^n \).

Thus,
\[
\mathbb{E}(|B_2|) \leq 2nd(1 - \varepsilon)^{d-1}.
\]
We prove that
\[
\mathbb{P}(|B_2| \geq 3nd(1 - \varepsilon)^{d-1}) = O(e^{-\Omega(n^{1/4})}).
\]
We note that changing a single random choice of a vertex in \( L \) can only change \( |B_2| \) by at most \( \Delta \). So we can use Lemma 4 once again with \( \mathcal{E} = \{ \Delta \leq n^{1/4} \} \), \( A_0 = n \), \( A_1 = n^{1/4} \) and \( t = nd(1 - \varepsilon)^{d-1} \).

**End of proof of Claim**

Given Claim 7, we have
\[
\mathbb{E}(\nu_{\ell}) = \mathbb{E}(\nu_{\ell} \mid \mathcal{B}) \mathbb{P}(\neg \mathcal{B}) + \mathbb{E}(\nu_{\ell} \mid \mathcal{B}) \mathbb{P}(\mathcal{B})
\]
\[
\leq n^\ell e^{-\varepsilon \ell^2/2m\ell^{-1}} \cdot \left( 1 + o(1) \right) \frac{d}{n} \ell^{-2} + O(n^\ell m^{\ell-1} e^{-\Omega(n^{1/4})}), \quad (16)
\]
\[
\leq n(1 + o(1))(d^2(1 - \varepsilon)^{-1} e^{-\varepsilon d^2/2})\ell^{-1} + o(1),
\]
\[
\leq ne^{-\varepsilon d^3/3},
\]
for \( d \) sufficiently large. This proves (14) with room to spare.

**Explanation of (16):** We can choose the vertex sequence \( \sigma = (x_1, y_1, \ldots, y_{\ell-1}, x_\ell) \) of an interesting path \( P \) in at most \( |B|^{\ell^2 m^{\ell-1}} \) ways, and we apply Claim 7. Having chosen \( \sigma \) we see that \((1 + o(1))d/n)^{2\ell-2}\) bounds the probability that the edges of \( P \) exist. To see this, condition on \( R_{n-1} \) and
the random choices for vertices not on \( P \). In particular, we can fix \( R_{n-1} \) from the beginning and this simply constrains the choices \( y_1, y_2, \ldots, y_{n-1} \) to be a random permutation of the chosen set. Let \( \mathcal{M} \) be the property that \( \Gamma \) has a matching from \( L \) to \( R \). It is known that \( P(\mathcal{M}) = 1 - O(n^{4-d}) \). This will also be true conditional on the value of \( R_{n-1} \). This follows by symmetry. The conditional spaces will be isomorphic to each other. So for large \( d \), we can assume that our conditioning is such that with probability \( 1 - O(1/n^2) \) the edge choices by \( x_1, x_2, \ldots, x_\ell \) are such that \( \Gamma \) has property \( \mathcal{M} \) with probability \( 1 - O(n^{7-d}) \). Recall from (2) that the disposition of the edges of \( \Gamma_{n-1} \) is independent of \( R_{n-1} \). Now each edge adjacent to a given \( x \in \sigma \cap L \) is a uniform choice over those edges consistent with \( x \) being in \( B \). But there will be \( n-1 \) such choices for such an \( x \) viz. the vertices of \( R_{n-1} \). Thus

\[
P(P \text{ exists } | \mathcal{M}) \leq \frac{P(P \text{ exists})}{P(\mathcal{M})} \leq (1 + o(1)) \left( \frac{d}{n-1} \right)^{2\ell-2}.
\]

Note that \( P(\bar{\mathcal{M}}) \) is only inflated by at most \( \frac{1}{(1-\varepsilon)^{2\ell}} = o(n^{o(1)}) \) if we condition on \( x_1, x_2, \ldots, x_\ell \) making their choices in \( R_{n-1} \). This has to be compared with the unconditional probability of \( O(n^{7-d}) \).

This completes the proof of Theorem 1.

Acknowledgement: We thank Wesley Pegden for his comments.

References


