On the insertion time of random walk cuckoo hashing

Alan Frieze∗ and Tony Johansson†
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh PA 15213
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 [8]). 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 [7] 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 \( R \). 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 the 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 \( \phi_k : L_k \rightarrow R_k \) 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 \( R_{k-1} = R \setminus R_{k-1} \).

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 \( N(v_1) \). Having defined \( M_k \) we proceed as follows: Steps 1 – 4 constitute round \( k \).

∗Research supported in part by NSF Grants DMS1362785, CCF1522984
†Research supported in part by NSF Grant DMS1362785
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** $M \leftarrow M \cup \{x, y\} \setminus \{y, \phi_{k-1}^{-1}(y)\}$; $x \leftarrow \phi_{k-1}^{-1}(y)$; 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 [6] gives a survey on Cuckoo Hashing and 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 $R$.

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,

$$E(|P_k|) \leq 2 \text{ for } k = 1, 2, \ldots, n. \quad (1)$$

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

The value 2 in (1) is the smallest whole number we can get from the analysis below. Replacing it by a larger constant relaxes the constraints on $d$.

## 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

$y_k$ is chosen uniformly from $\bar{R}_{k-1}$ \hspace{1cm} (2)

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 anymore 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 $R$.

Let

$$B = \{v \in L : N(v) \cap \bar{R}_{n-1} = \emptyset\}.$$
If $x \not\in B$ in Step 2 of \textsc{insert} then we will have found $P_n$.

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

Let $\nu_{\ell}$ denote the number of interesting paths with $2\ell - 1$ vertices.

\textbf{Lemma 2} \textit{Given $A_0$ and $d$ sufficiently large,}

\begin{equation}
\mathbb{P}(2 \leq \ell \leq A_0 \log \log n : \nu_{\ell} \geq ne^{-\varepsilon \ell d/3}) = o(n^{-2}).
\end{equation}

\textit{Here our probability is over $\Gamma$ and the choice of $R_{n-1}$}.

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

\textbf{Claim 3} \textit{W.h.p. $\Gamma$ contains at most $n^{1/2}$ cycles of length at most $\lambda_d = (\log_d n)^{1/100}$}.

\textbf{Proof of Claim:} The expected number of cycles of length at most $2\ell = \lambda_d$ is bounded by

\begin{equation}
\binom{n}{\ell} \binom{m}{\ell} (\ell!)^2 \left( \frac{d}{m} \right)^{2\ell} \leq (de)^{2\ell} \leq n^{1/50}.
\end{equation}

The claim follows from the Markov inequality.

\textit{End of proof of Claim}

\textbf{Claim 4} \textit{W.h.p. $\Gamma$ has maximum degree at most $\log n$.}

\textbf{Proof of Claim:} If $v \in L$ then its degree $d(v) = d$. Now consider $w \in R$. Then for $\ell = \log n$,

\begin{equation}
\mathbb{P}(\exists w \in R : d(w) \geq \ell) \leq m \left( \frac{dn}{\ell} \right)^{\frac{1}{m^2}} \leq m \left( \frac{de}{\ell} \right)^{\ell} \leq n^{- \log \log n / 2}.
\end{equation}

\textit{End of proof of Claim}

As a corollary of Claim 3 and Claim 4 we see that if $2\ell = (\log n)^{1/100}$,

\begin{equation}
\text{w.h.p. there are at most } n^{1/2} (\log n)^{2\ell} = n^{1/2 + o(1)} \text{ vertices within distance } 2\ell \text{ of a cycle of length at most } 2\ell.
\end{equation}

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.

\textbf{Claim 5} There exists a constant $a > 0$ such that for any $v \in L_{n-1}$, the expected time for \textsc{insert} to reach $\overline{R}_{n-1}$ is $O((\log n)^a)$.
Now let $p_\ell$ denote the probability that INSERT requires at least $\ell$ rounds to insert $v_{n-1}$. We prove the theorem by showing that

$$E(|P_n|) = 1 + 2 \sum_{\ell=2}^{\infty} p_\ell \leq 2. \quad (5)$$

(We only need to verify the second equation.)

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

$$A_0 \log \log n \sum_{\ell=2}^{A_0 \log \log n} p_\ell \leq O(n^{-1/2+o(1)}) + \sum_{\ell=2}^{A_0 \log \log n} \frac{\nu_\ell}{n(d-1)^\ell} \leq O(n^{-1/2+o(1)}) + \frac{ne^{-\epsilon d/3}}{n(d-1)^\ell} \leq \frac{1}{3}. \quad (6)$$

**Explanation of (6):** Following (4), we find that the probability $v_n$ is within $2A_0 \log \log n$ of a cycle of length at most $\lambda_d$ 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. Failing this, we have divided the number of interesting paths of length $2\ell$ 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 $y \in \bar{R}_{n-1}$ then instead of stopping, we move to $v_n$ and continue. In this way there 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 $\bar{R}_{n-1}$. Finally, for the algorithm to take at least $\ell$ rounds, it must choose an interesting path.

Note next that

$$p_{A_0 \log \log n} \leq O(n^{-1/2+o(1)}) + \frac{e^{-\epsilon d} A_0 \log \log n/3}{(d-1)A_0 \log \log n} \leq \frac{1}{3(\log n)^{A_1}},$$

where the constant $A_1$ can be made as large as necessary by increasing $d$.

It follows that

$$\sum_{\ell=A_0 \log \log n}^{(\log n)^{A_1}} p_\ell \leq \sum_{\ell=A_0 \log \log n}^{(\log n)^{A_1}} p_{A_0 \log \log n} \leq \frac{1}{3}. \quad (7)$$

It follows from Claim 5 that for any integer $\rho \geq 1$,

$$P(|P_n| \geq \rho (\log n)^{2a}) \leq \frac{1}{(\log n)^{\rho a}}. \quad (8)$$

It follows from (8) that

$$\sum_{\ell \geq (\log n)^{2a}} p_\ell \leq \sum_{\rho=1}^{\infty} \sum_{\ell/(\log n)^{2a} \in [\rho, \rho+1]} p_\ell \leq \sum_{\rho=1}^{\infty} \frac{1}{(\log n)^{\rho a-2}} = o(1). \quad (9)$$

Theorem 1 now follows from (5), (6), (7) and (9), if we take $A_1 \geq 2a$.

### 2.1 Proof of Lemma 2

Fix $2 \leq \ell \leq A_0 \log \log n$ and let $P_\ell$ denote the set of interesting paths of length $2\ell - 1$. For $P \in P_\ell$ we let $d(P)$ denote the number of $Q \in P_\ell$ such that $P, Q$ share a vertex. Then we let
\( \Delta_t = \max_{P \in \mathcal{P}_t} d(P) \). The Markov inequality implies that for any \( \alpha > 0 \) and \( t = 1, 2, \ldots \),

\[
P(\nu_t \geq \alpha n) = P((\nu_t)_t \geq (\alpha n)_t) \leq \frac{E((\nu_t)_t)}{(\alpha n)_t}. \tag{10}
\]

Here \( (m)_t = m(m - 1) \cdots (m - t + 1) \).

We will use the inequality

\[
E((\nu_\ell)_t) \leq \sum_{i=0}^{t-1} (\Delta (t - i))^i E(\nu_\ell)^{t-i}. \tag{11}
\]

The expectation on the LHS of (11) is of the number of distinct \( t \)-sequences of paths in \( \mathcal{P}_t \). A summand bounds the expected number of sequences obtained by choosing \( t - i \) vertex disjoint paths and then choosing \( i \) paths that meet one of these \( t - i \) paths. The lemma will follow from

\[
P(\Delta \geq (\log n)^2) \leq n^{-\omega} \text{ where } \omega \to \infty. \tag{12}
\]

\[
E(\nu_\ell) \leq \beta n \text{ where } \beta = e^{-\varepsilon \ell d/3}. \tag{13}
\]

Indeed, substituting (12), (13) into (11) gives that for \( t = (\log n)^2 \),

\[
E((\nu_\ell)_t) \leq \sum_{i=0}^{t-1} t^i (\log n)^{ti}(\beta n)^{t-i} \leq (2\beta n)^t. \tag{14}
\]

Substituting (14) into (10) gives, for \( t = (\log n)^2 \) and \( \alpha = 12\beta \),

\[
P(\nu_t \geq \alpha n) \leq \frac{(2ne^{-\varepsilon \ell d/3})^t}{(\alpha n)^t} \leq \left( \frac{2\varepsilon \beta}{\alpha} \right)^t \leq 2^{-t}.
\]

**Proof of (12)**

It follows from (3) that with the claimed probability,

\[
\Delta \leq 2\ell d^\ell (\log n)^\ell \leq (\log n)^{2\ell}.
\]

**Proof of (13)**

**Claim 6** Let

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

Then

\[
P(\mathcal{B}) = O(n^{-2}).
\]

**Proof of Claim:** First of all let

\[
K = \{ k : \text{round } k \text{ does not end immediately in Step 2 with } x = v_k. \}
\]

Then, \( P(k \in K) \leq \left( \frac{k}{m} \right)^d \) and this holds for each value of \( k \) independently and so

\[
E(|K|) \leq \sum_{k=1}^n \left( \frac{k}{m} \right)^d \leq \frac{n^{d+1}}{dm^d} \leq nd^{-1}e^{-\varepsilon d}.
\]
Now $|K|$ is the sum of independent $\{0,1\}$ random variables and so Hoeffding’s theorem implies that
\[
\Pr(|K| \geq 2nd^{-1}e^{-\varepsilon d}) = O(n^{-2}),
\]
with room to spare.

Now if $B_1 = \{ v_k \in B : \exists \ell \neq k \text{ s.t. round } \ell \text{ ends with } x = v_k \}$ then $|B_1| \leq |K|$. But if $B_2 = B \setminus B_1$ then $v_k \in B_2$ only if random choices of other vertices of $L$ include all the neighbors of $v_k$. Then, it follows from (2) that
\[
\Pr(v_k \in B_2) \leq \left(1 - \left(1 - \frac{1}{m}\right)^n\right)^d \leq (1 - \varepsilon)^d.
\]

It is straightforward to show concentration of $|B_2|$ using the Azuma-Hoeffding inequality.

**End of proof of Claim**

Given Claim 6, we have
\[
E(\nu_\ell) = E(\nu_\ell | -B)\Pr(-B) + E(\nu_\ell | B)\Pr(B)
\leq n^\ell e^{-\varepsilon d/2}m^{\ell-1} \cdot (1 + o(1))\left(\frac{d}{n-d}\right)^{2\ell-2} + O(n\Delta^\ell \cdot n^{-2}),
\]
\[
\leq n(1 + o(1))(d^2(1 - \varepsilon)^{-1}e^{-\varepsilon d/2})^{\ell} + o(1),
\leq ne^{-\varepsilon d/3},
\]
for $d$ sufficiently large.

**Explanation of (15):** We choose the vertex sequence $\sigma = (x_1, y_1, \ldots, y_{\ell-1}, x_\ell)$ of an interesting path $P$ in at most $|B|^\ell m^{\ell-1}$ ways. Having chosen $\sigma$ we see that $((1 + o(1))d/(n-1))^{2\ell-2}$ 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 $P$. Let $\mathcal{M}$ be the property that $\Gamma$ has a matching from $L$ to $R$. It is known that $\Pr(\mathcal{M}) = 1 - O(n^{4-d})$. 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 almost all edge choices by $x_1, x_2, \ldots, x_\ell$ are such that $\Gamma$ has property $\mathcal{M}$. 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 $x \in P \cap L$ is a uniform choice over those edges consistent with $x$ being in $B$. But there will always be at least $n-1$ such choices for such an $x$ viz. the vertices of $R_{n-1}$. Thus
\[
\Pr(P | \mathcal{M}) \leq \frac{\Pr(P)}{\Pr(\mathcal{M})} \leq (1 + o(1))\frac{(n-1)(n-2)^{\ell-2}(n-1)^{\ell}}{(n-1)^{\ell}} \leq (1 + o(1))\left(\frac{d}{n-d}\right)^{2\ell-2}.
\]

Note that $\Pr(\bar{\mathcal{M}})$ is only inflated by at most $(1 - \varepsilon)^{-d\ell}$ if we condition on $x_1, x_2, \ldots, x_\ell$ making their choices in $R_{n-1}$. Thus $\Pr(\bar{\mathcal{M}}) = O(n^{4-d}(1 - \varepsilon)^{-d\ell}) = o(1)$, still.

**Acknowledgement:** We thank Wesley Pegden for his comments.

**References**


