Sharded Shuffle | DRAFT 2024-03-12

Erik Taubeneck

Context

We propose a shuffle mechanism, $\pi(x)$, which can operated in a distributed fashion, called a sharded shuffle. Given S shards, and a set $\mathbb{X} = \{x\}$ of cardinality n this operates by:

For each $x \in \mathbb{X}$, compute $s(x) = rand(0, S-1)$, where $rand(0, S-1)$ picks a value in $[0, S)$ with uniform probability, and place x on the $s(x)$th machine.

For each machine $m \in [0, S-1]$, perform a uniform random shuffle, $\underset{m}{\pi}(x)$, of $\{x \in \mathbb{X} | s(x) = m\}$.

We can think of $\pi(x)$ as composition of $s(x) \circ \pi_{s(x)}(x)$.

Let $\mathbb{K} = \{k_i = |\{x \in \mathbb{X} | s(x) = m\} | \forall i \in [0, S-1]\}$, e.g. the size of each shard. Note that $\sum k_i = n$.

We define the order of the resulting process by iterating through the items on each shard, shard by shard. That is, the index i of the result of $\pi(x)$ is $\left(\sum_{j=0}^{m-1} k_j\right) + \underset{m}{\pi}(x)$

We aim to show that this process produces a uniform random permutation, and that, even with knowledge of $\mathbb{K}$, so long as the assignment $s(x)$ and the per-shard uniform shuffle are oblivious, the entire process is oblivious.

Proof

Lemma 1. *Given some $\mathbb{K}$ (e.g., the size of the shards), the number of permutations is $n!$.*

Proof. First, define $r_0 = n$ and $r_i = r_{i-1} - k_{i-1}$, for $i \in [1, S)$. If we go shard by shard, this is the number of remaining values for all shards $m \geq i$.

Now, note that the number of permutations can be expressed as:

$$\binom{n}{k_0} k_0! \cdot \binom{r_1}{k_1} k_1! \cdots \binom{r_{S-1}}{k_{S-1}} k_{S-1}!.$$

That is, at for each shard m, choose k_m elements from the remaining r_m elements. For each of those shard configurations, they can then be shuffled $k_m!$ ways.

Rewriting, and replacing $r_0 = n$:

$$\prod_{i=0}^{S-1} \binom{r_i}{k_i} k_i! = \prod_{i=0}^{S-1} \frac{r_i!}{k_i!(r_i - k_i)!} \cdot k_i! = \prod_{i=0}^{S-1} \frac{r_i!}{(r_i - k_i)!}.$$

Now, we can rearrange this, by shifting the numerators one to the left, e.g.

$$= r_0! \cdot \left(\prod_{i=1}^{S-1} \frac{r_i!}{(r_{i-1} - k_{i-1})!} \right) \cdot \frac{1}{(r_{S-1} - k_{S-1})!}.$$

By construction, $r_i = r_{i-1} - k_{i-1}$. Replacing the denominator in the product term, we get

$$= r_0! \cdot \left(\prod_{i=1}^{S-1} \frac{r_i!}{r_i!} \right) \cdot \frac{1}{(r_{S-1} - k_{S-1})!} = r_0! \cdot \left(\prod_{i=1}^{S-1} 1 \right) \cdot \frac{1}{(r_{S-1} - k_{S-1})!} = r_0! \cdot \frac{1}{(r_{S-1} - k_{S-1})!}.$$

Now, note that $r_i = n - \sum_{j=0}^{i-1} k_j$, as it's simply repeated subtraction of the next k_j. Thus,

$$r_{S-1} - k_{S-1} = n - \left(\sum_{j=0}^{(S-1)-1} k_j \right) - k_{S-1} = n - \sum_{j=0}^{(S-1)} k_j.$$

By construction, this sum is n, and thus

$$r_{S-1} - k_{S-1} = n - \sum_{j=0}^{S-1} k_j = n - n = 0.$$

Therefore, our original number of permutations

$$\prod_{i=0}^{S-1} \binom{r_i}{k_i} k_i! = r_0! \cdot \frac{1}{(r_{S-1} - k_{S-1})!} = r_0! \cdot \frac{1}{0!} = r_0! = n!.$$

□

Lemma 2. *Given some $\mathbb{K}$, for all $x \in \mathbb{X}$, $y \in \mathbb{X}$, $P(\pi(x) = y) = 1/n$.*

Proof. Let $s(x)$ be the shard assignment of x. Note that because $\mathbb{K}$ is given,

$$P(s(x) = i) = \frac{k_i}{n}.$$

Also note that because π uses a uniform random shuffle on each shard,

$$P(\pi(x) = y | k(x) = i \wedge \sum_{j=0}^{i-1} k_j < y \leq \sum_{j=0}^{i} k_j) = \frac{1}{k_i}.$$

e.g., the given x lands on the same shard as y, the probability of landing on y is $\frac{1}{k_i}$, 1 over the size of that shard.

Then

$$P(\pi(x) = y) = P(k(x) = i \wedge \sum_{j=0}^{i-1} k_j < y \leq \sum_{j=0}^{i} k_j) \cdot P(\pi(x) = y | k(x) = i \wedge \sum_{j=0}^{i-1} k_j < y \leq \sum_{j=0}^{i} k_j) = \frac{k_i}{n} \frac{1}{k_i} = \frac{1}{n}.$$

□

Lemma 3. *Given some $\mathbb{K}$, and any permutation π_i given $\mathbb{K}$, $P(\pi = \pi_i) = \frac{1}{n!}$*

Proof. Note that we can expand $P(\pi = \pi_i)$ to

$$P(\pi = \pi_i) = P\left(\bigcap_{x=0}^{n} \pi(x) = \pi_i(x)\right) = \prod_{x=0}^{n} P\Big(\pi(x) = \pi_i(x) | \pi(\xi) = \pi_i(\xi) \forall \xi \in [0, x)\Big).$$

By Lemma 2, for $x = 0$:

$$P\Big(\pi(0) = \pi_i(0) | \pi(\xi) = \pi_i(\xi) \forall \xi \in [0, 0)\Big) = P(\pi(0) = \pi_i(0)) = \frac{1}{n}.$$

Then, for $x \in [1, n)$, we can think of this as a new set with $n - x$ elements, with a permutation $\pi^{(n-x)}$.

$\mathbb{K}_x$ becomes $\{k_i - |\{s(\xi) = i \forall \xi \in [0, x)\}| \forall k_i \in \mathbb{K}\}$, that is the original shard sizes minus the values chosen up to x.

$$P\Big(\pi(x) = \pi_i(x) | \pi(\xi) = \pi_i(\xi) \forall \xi \in [0, x)\Big) = P\Big(\pi^{(n-x)}(x) = \pi_i(x)\Big) + \sum_{\xi=0}^{x} P(\pi_i(\xi) \in \pi_i(\hat{\xi}) \forall \hat{\xi} \in [0, x)]).$$

Since all the values $\pi_i(\hat{\xi}) \forall \hat{\xi} \in [0, x)]$ have all been selected, every term in the summation is 0. We can again apply Lemma 2:

$$P\Big(\pi^{(n-x)}(x) = \pi_i(x)\Big) = \frac{1}{n - x}.$$

Plugging this into our original equation:

$$P(\pi = \pi_i) = \prod_{x=0}^{n} P\Big(\pi(x) = \pi_i(x) | \pi(\xi) = \pi_i(\xi) \forall \xi \in [0, x)\Big) = \prod_{x=0}^{n} \frac{1}{n - x} = \frac{1}{n!}$$

□

Now, because of Lemma 1, we know that there are $n!$ possible permutations, and by Lemma 3, each have equal probability $\frac{1}{n!}$. Thus, for a given $\mathbb{K}$, π is a uniform random permutation and no biased is introduced when $\mathbb{K}$ is known.

For the final step, we just need to show that π is in general a random permutation. Let κ be all possible $\mathbb{K}$. Then, for any $\mathbb{K}_i \in \kappa$, there exists some probability, $p_i \in [0, 1]$, of occurring when performing π. Now, by Lemma 3, we know that for a given $\mathbb{K}_i \in \kappa$, $P(\pi = \pi_j) = \frac{1}{n!}$. Thus, across all $\mathbb{K}_i \in \kappa$ we can compute

$$\sum_{\mathbb{K}_i \in \kappa} p_i \cdot P(\pi = \pi_j | \mathbb{K}_i) = \sum_{\mathbb{K}_i \in \kappa} p_i \frac{1}{n!} = \frac{1}{n!} \sum_{\mathbb{K}_i \in \kappa} p_i = \frac{1}{n!}.$$

$\mathbb{QED}$.