# Supplementary Material

## A TrieHH for the Setting of Multiple Sequences per User

**Algorithm 3** A Trie-based Frequent Sequence Algorithm $\mathcal{M}(\mathcal{D}, \theta, \gamma)$ for Multiple Sequences per User.

```
Input: A set $\mathcal{D} = \{u_1, u_2, \ldots, u_n\}$, A threshold $\theta$. Batch size $m = \gamma\sqrt{n}$.
Output: A trie.
Set $T = \{root\}$; $T_{old} = None$; $i = 1$;
while $T! = T_{old}$ do
    Choose $m$ users from $\mathcal{D}$ uniformly at random, denote as $\tilde{\mathcal{X}}$. Initialize $\tilde{\mathcal{X}} = \{\}$.
    for For each user $u_i \in \tilde{\mathcal{X}}$ do
        Randomly select a sequence $w_j \in u_i$ with respect to its frequency $f_i(w_j)$ in $u_i$, and add $w_j$ to $\tilde{\mathcal{X}}$.
    end for
    $T_{old} = T$;
    $T = \mathcal{V}(\tilde{\mathcal{X}}, T, \theta, i)$;   $i$++;
end while
return $T$;
```

## B Proof of Theorem 1 and Theorem 2

We will show that when $n \geq 10^4$, choosing $\theta \geq 10$, $\gamma \geq 1$, and $\gamma \leq \frac{\sqrt{n}}{\theta+1}$, ensures that Algorithm 1 is $\left(L \ln\left(1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1}\right), \frac{\theta-2}{(\theta-3)\theta!}\right)$-differentially private. This theorem is proved by combining two lemmas that deal with different cases of the population. In Lemma 1, we first show a bound on the ratio between $P(\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T_{i-1}) = T_i)$ and $P(\mathcal{M}_i(\mathcal{D}', \theta, \gamma, T_{i-1}) = T_i)$ for any trie $T \in Range(\mathcal{M})$ that $p_i(w) \in T_i$ . This bound depends on $k$, the number of sequences that have prefix $p_i(w)$ in $\mathcal{D}'$. It is obvious that when $k = \theta - 1$, $P(\mathcal{M}_i(\mathcal{D}', \theta, \gamma, T_{i-1}) = T_i)$ must be 0, but the number of sequences having prefix $p_i(w)$ in $\mathcal{D}$ is $\theta$, so $P(\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T_{i-1}) = T_i)$ is greater than 0. In this case, the ratio between them approaches infinity. On the one hand, if the number of sequences with prefix $p_i(w)$ in $\mathcal{D}'$ is already large, then an extra $p_i(w)$ in $\mathcal{D}$ only affects the probability slightly, so the ratio between $P(\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T_{i-1}) = T_i)$ and $P(\mathcal{M}_i(\mathcal{D}', \theta, \gamma, T_{i-1}) = T_i)$ is small, and it could be bounded by a small $\varepsilon$. On the other hand, if the number of sequences with $p_i(w)$ in $\mathcal{D}$ is actually small, then the probability $P(\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T_{i-1}) = T_i)$ is small, and could be bounded by a reasonably small $\delta$. This case is handled by Lemma 2.

We start by calculating the probability that a prefix $p$ appears at least $\theta$ times if we randomly choose $m$ users from a pool of users of size $n$, assuming that $p$ appears $W$ times in the population.

**Proposition 2.** *Suppose prefix $p$ appears $W$ times in a pool of $n$ users. If we select $m$ users uniformly at random from them, then the probability that prefix $p$ is appears at least $\theta$ times is*

$$\frac{1}{\binom{n}{m}} \sum_{i=\theta}^{\min\{W,m\}} \binom{W}{i} \binom{n-W}{m-i}$$

*Proof.* The probability that a prefix $p$ appears $i$ times follows the hypergeometric distribution $P(i) = \frac{1}{\binom{n}{m}}\binom{W}{i}\binom{n-W}{m-i}$. To calculate the probability that $p$ appears at least $\theta$ times in the chosen subset, we sum up the case that $p$ appears $\theta, \theta+1, \ldots, \min\{W, m\}$ times. □

The above probability expression will be useful in the proof of Lemma 2 below, and when we investigate the privacy-utility trade-off in Section 3. Also, Proposition 1 is derived from Proposition 2.

**Lemma 1.** *$\forall T \in Range(\mathcal{M})$ such that $p_i(w) \in T_i$, $\forall i \in \{1, \ldots, l\}$, assume there are $k$ users in $\mathcal{D}'$ that have prefix $p_i(w)$, and $k \geq \theta$. Then $P(\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T_{i-1}) = T_i) \leq (1 + \frac{\theta}{k-\theta+1})P(\mathcal{M}_i(\mathcal{D}', \theta, \gamma, T_{i-1}) = T_i)$.*

*Proof.* Let $C(\mathcal{D}, \mu, \theta, \gamma, T_{in}, T_{out})$ be a function to count the number of ways to choose $\mu$ users (denote the set of chosen users as $\tilde{\mathcal{X}}$) from a set of users $\mathcal{D}$, that using Algorithm 2, $\mathcal{V}(\tilde{\mathcal{X}}, T_{in}, \theta, i) = T_{out}$. Also, we denote $C(\mathcal{D}, \mu, \theta, \gamma, T_{in}, T_{out}|p_i(w))$ as the number of ways to choose users under the same condition, given prefix $p_i(w)$ is added to $T_{out}$ in this step.

Remember $\mathcal{D}$ and $\mathcal{D}'$ differ in only one sequence $w$ and $w'$, that $|w| = l$, $|w'| = 0$. We denote $w$'s prefix of length $i$ as $p_i(w)$. For any output trie $T \in \mathit{Range}(\mathcal{M})$, consider the step to grow $T_i$ from $T_{i-1}$ by $M_i$. Let $\mathcal{Z} = \mathcal{D} - \{w\} = \mathcal{D}' - \{w'\}$. We assume there are $k$ users in $\mathcal{D}'$ that have prefix $p_i(w)$ of $w$. We denote this subset of users in $\mathcal{D}'$ as $\mathcal{W}$. Thus the set of users in $\mathcal{D}$ that have prefix $p_i(w)$ is $\mathcal{W} + \{w\}$ with size $k + 1$.

We abuse the notation to use $C(\mathcal{D}, \mu)$ instead of $C(\mathcal{D}, \mu, \theta, \gamma, T_{in}, T_{out})$, and $C(\mathcal{D}, \mu|p_i(w))$ instead of $C(\mathcal{D}, \mu, \theta, \gamma, T_{in}, T_{out}|p_i(w))$ for fixed $\theta$, $\gamma$, $T_{in}$, $T_{out}$.

We calculate $P(\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T_{i-1}) = T_i)$ by the ratio between $C(\mathcal{D}, m, \theta, \gamma, T_{i-1}, T_i)$ (how many ways to choose $m$ users from $\mathcal{D}$, that $\mathcal{V}(\mathcal{D}, T_{in}, \theta, i)$ returns $T_{out}$ and $\binom{n}{m}$ (how many ways to choose $m$ users from $\mathcal{D}$).

Also, we could separate $C(\mathcal{D}', m, \theta, \gamma, T_{i-1}, T_i)$ into two parts: not choosing $w'$ (taking all $m$ users from $\mathcal{D}' - \{w'\}$), or choosing $w'$ (taking the rest $m - 1$ users from $\mathcal{D}' - \{w'\}$). Thus,

$$\begin{aligned} P(\mathcal{M}_i(\mathcal{D}', \theta, \gamma, T_{i-1}) = T_i) &= \frac{C(\mathcal{D}', m, \theta, \gamma, T_{i-1}, T_i)}{\binom{n}{m}} \\ &= \frac{1}{\binom{n}{m}}(C(\mathcal{D}' - \{w'\}, m) + C(\mathcal{D}' - \{w'\}, m-1)) \\ &= \frac{1}{\binom{n}{m}}(C(\mathcal{Z}, m) + C(\mathcal{Z}, m-1)) \end{aligned} \tag{3}$$

Consider $C(\mathcal{Z}, m-1, \theta, \gamma, T_{i-1}, T_i)$, because $p_i(w) \in T_i$, so there must be at least $\theta$ users in the chosen set voting for $p_i(w)$. We consider the following cases separately: choosing $\theta$ users from $\mathcal{W}$, $m - 1 - \theta$ users from $\mathcal{Z} - \mathcal{W}$ (note that $p_i(w) \in T_i$ is already guaranteed by choosing $\theta$ users from $\mathcal{W}$, so we consider $p_i(w)$ as a given condition here), and choosing $\theta + 1$ users from $\mathcal{W}$, $m - \theta - 1$ users from $\mathcal{Z} - \mathcal{W}$, ..., i.e.,

$$C(\mathcal{Z}, m-1) = \sum_{i=\theta}^{\min\{k,m\}} \binom{k}{i} C(\mathcal{Z} - \mathcal{W}, m - i - 1|p_i(w))$$

Similarly for $C(\mathcal{Z}, m)$, not choosing $w$ (taking all $m$ users from $\mathcal{D} - \{w\}$) or choosing $w$ (taking the rest $m - 1$ users from $\mathcal{D}' - \{w\}$).

$$\begin{aligned} C(\mathcal{Z}, m) = &\binom{k}{\theta} C(\mathcal{Z} - \mathcal{W}, m - \theta|p_i(w)) \\ &+ \sum_{i=\theta+1}^{\min\{k,m\}} \binom{k}{i} C(\mathcal{Z} - \mathcal{W}, m - i|p_i(w)) \end{aligned}$$

Thus,

$$C(\mathcal{Z} - \mathcal{W}, m - \theta|p_i(w)) \leq \frac{1}{\binom{k}{\theta}} C(\mathcal{Z}, m) \tag{4}$$

$C(\mathcal{D}, m, \theta, \gamma, T_{i-1}, T_i)$ could also be considered as not choosing $w$ (taking all $m$ users from $\mathcal{D} - \{w\}$) or choosing $w$ (taking the rest $m - 1$ users from $\mathcal{D} - \{w\}$). But different from $\mathcal{D}'$, if $w \in \mathcal{D}$ is chosen, we can choose $\theta - 1$ to $k$ users contain prefix $p_i(w)$ from $\mathcal{Z} - \mathcal{W}$. Thus,

$$
\begin{aligned}
&C(\mathcal{D}, m, \theta, \gamma, T_{i-1}, T_i)\\
&= C(\mathcal{Z}, m) + \binom{k}{\theta-1} C(\mathcal{Z}-\mathcal{W}, m-\theta | p_i(w))\\
&+ \sum_{i=\theta}^{\min\{k,m\}} \binom{k}{i} C(\mathcal{Z}-\mathcal{W}, m-i-1 | p_i(w))\\
&= C(\mathcal{Z}, m) + \binom{k}{\theta-1} C(\mathcal{Z}-\mathcal{W}, m-\theta | p_i(w))\\
&+ C(\mathcal{Z}, m-1) \qquad (5)
\end{aligned}
$$

By Equation 5 and Inequality 4,

$$
\begin{aligned}
&P(\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T_{i-1}) = T_i) = \frac{C(\mathcal{D}, m, \theta, \gamma, T_{i-1}, T_i)}{\binom{n}{m}}\\
&= \frac{1}{\binom{n}{m}} (C(\mathcal{Z}, m) + \binom{k}{\theta-1} C(\mathcal{Z}-\mathcal{W}, m-\theta | p_i(w))\\
&+ C(\mathcal{Z}, m-1))\\
&\leq \frac{1}{\binom{n}{m}} (C(\mathcal{Z}, m) + C(\mathcal{Z}, m-1) + \frac{\binom{k}{\theta-1}}{\binom{k}{\theta}} C(\mathcal{Z}, m))\\
&= \frac{1}{\binom{n}{m}} (C(\mathcal{Z}, m) + C(\mathcal{Z}, m-1)\\
&+ \frac{\theta}{k-\theta+1} C(\mathcal{Z}, m))\\
&\leq (1 + \frac{\theta}{k-\theta+1}) \frac{1}{\binom{n}{m}} (C(\mathcal{Z}, m) + C(\mathcal{Z}, m-1))\\
&= (1 + \frac{\theta}{k-\theta+1}) P(\mathcal{M}_i(\mathcal{D}', \theta, \gamma, T_{i-1}) = T_i)
\end{aligned}
$$

□

Suppose there are $k$ users has prefix $p_1(w)$. In Lemma 2, we show that when $k \leq \frac{\sqrt{n}}{\gamma} - 1$, $P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma)) \leq \frac{\theta-2}{(\theta-3)\theta!}$. This means when $k$ is small, the probability that $p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma)$ is small, so it could be bounded by a small $\delta$. And it is the same for the $i^th$ round that when there are $k$ users has prefix $p_i(w)$. If $k \leq \frac{\sqrt{n}}{\gamma} - 1$, then $P(p_{i-1}(w) \in \mathcal{M}(\mathcal{D}, \theta, \gamma) | p_{i-2}(w) \in \mathcal{M}(\mathcal{D}, \theta, \gamma)) \leq \frac{\theta-2}{(\theta-3)\theta!}$.

**Lemma 2.** *Consider the step to grow $T_i$ from $T_{i-1}$ by $\mathcal{M}_i$. We assume there are $k$ users in $\mathcal{D}'$ that have prefix $p_i(w)$ of $w$. Then there are $k+1$ users in $\mathcal{D}$ that have prefix $p_i(w)$. When $k \leq \frac{\sqrt{n}}{\gamma} - 1$, $4 \leq \theta \leq \sqrt{n}$, $\gamma \geq 1$, $P(p_i(w) \in \mathcal{M}(\mathcal{D}, \theta, \gamma) | p_{i-1}(w) \in \mathcal{M}(\mathcal{D}, \theta, \gamma)) \leq \frac{\theta-2}{(\theta-3)\theta!}$. For the first step, $P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma)) \leq \frac{\theta-2}{(\theta-3)\theta!}$.*

*Proof.* First $P(p_1(w) \in \mathcal{M}(\mathcal{D}, \theta, \gamma)) \leq P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma, T_0))$. To calculate $P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma, T_0))$, we consider the cases of choosing $\theta$ to $k+1$ users voting for $p_i(w)$ separately, By Proposition 2,

$$
\begin{aligned}
&P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma, T_0))\\
&= \frac{1}{\binom{n}{m}} \sum_{i=\theta}^{\min\{k+1,m\}} \binom{k+1}{i} \binom{n-k-1}{m-i}
\end{aligned}
$$

Note that when $k+1<\theta$, $P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma, T_0)) = 0$, so we only consider the case that $k+1 \geq \theta$. The sum of the array above could be upper bounded by the sum of a geometric sequence. We know that $k \leq \frac{\sqrt{n}}{\gamma} - 1$, $m = \gamma\sqrt{n}$. Consider the ratio between the first two items,

$$\begin{aligned}
\frac{\binom{k+1}{\theta+1}\binom{n-k-1}{m-\theta-1}}{\binom{k+1}{\theta}\binom{n-k-1}{m-\theta}} &= \frac{(k-\theta+1)(m-\theta)}{(\theta+1)(n-k-m+\theta)} \\
&\leq \frac{(\frac{\sqrt{n}}{\gamma}-\theta)(\gamma\sqrt{n}-\theta)}{(\theta+1)(n-\frac{\sqrt{n}}{\gamma}+1-\gamma\sqrt{n}+\theta)} \\
&\leq \frac{n}{(\theta+1)(n-(\gamma+\frac{1}{\gamma})\sqrt{n}+1+\theta)} \\
&\leq \frac{1}{(\theta+1)(1-\frac{\gamma+\frac{1}{\gamma}}{\sqrt{n}})} \qquad (6)
\end{aligned}$$

We denote $\frac{1}{(\theta+1)(1-\frac{\gamma+\frac{1}{\gamma}}{\sqrt{n}})}$ as $r_1$. Because $k \leq \frac{\sqrt{n}}{\gamma} - 1$ and $k+1 \geq \theta$, so $\gamma \leq \frac{\sqrt{n}}{k+1} \leq \frac{\sqrt{n}}{\theta}$. We know that $\gamma \geq 1$, then $\gamma + \frac{1}{\gamma} \leq \gamma + 1 \leq \frac{\sqrt{n}}{\theta} + 1$. And also $\theta \leq \sqrt{n}$, so $\gamma + \frac{1}{\gamma} \leq \frac{\sqrt{n}}{\theta} + 1 \leq 2\frac{\sqrt{n}}{\theta}$. Now we are able to bound $r_1$:

$$\begin{aligned}
r_1 &= \frac{1}{(\theta+1)(1-\frac{\gamma+\frac{1}{\gamma}}{\sqrt{n}})} \\
&\leq \frac{1}{(\theta+1)(1-\frac{\frac{2\sqrt{n}}{\theta}}{\sqrt{n}})} \\
&\leq \frac{1}{(\theta+1)(1-\frac{2}{\theta})} \\
&\leq \frac{1}{\theta(1-\frac{2}{\theta})} \\
&= \frac{1}{\theta-2}
\end{aligned}$$

Note that when $\theta \geq 4$, $r_1 < 1$.

Now we'll show that the ratio between adjacent items is decreasing. Consider the ratio between any two adjacent items $\binom{k+1}{\theta+i+1}\binom{n-k-1}{m-\theta-i-1}$ and $\binom{k+1}{\theta+i}\binom{n-k-1}{m-\theta-i}$,

$$\begin{aligned}
\frac{\binom{k+1}{\theta+i+1}\binom{n-k-1}{m-\theta-i-1}}{\binom{k+1}{\theta+i}\binom{n-k-1}{m-\theta-i}} &= \frac{(k-\theta-i+1)(m-\theta-i)}{(\theta+i+1)(n-k-m+\theta+i)} \\
&\leq \frac{(k-\theta+1)(m-\theta)}{(\theta+1)(n-k-m+\theta)} \\
&= \frac{\binom{k+1}{\theta+1}\binom{n-k-1}{m-\theta-1}}{\binom{k+1}{\theta}\binom{n-k-1}{m-\theta}} \\
&\leq r_1
\end{aligned}$$

Thus,

$$\begin{aligned}
&P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma, T_0)) \\
&= \frac{1}{\binom{n}{m}} \sum_{i=\theta}^{k+1} \binom{k+1}{i} \binom{n-k-1}{m-i} \\
&\leq \frac{1}{\binom{n}{m}} (\sum_{i=0}^{k+1-\theta} r_1^i) \binom{k+1}{\theta} \binom{n-k-1}{m-\theta} \\
&\leq \frac{1}{\binom{n}{m}} \frac{1}{1-r_1} \binom{k+1}{\theta} \binom{n-k-1}{m-\theta}
\end{aligned}$$

The last line follows because $r_1 < 1$.

When $k \leq \frac{\sqrt{n}}{\gamma} - 1$,

$$\begin{aligned}
&P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma, T_0)) \\
&\leq \frac{1}{\binom{n}{m}} \frac{1}{1-r_1} \binom{k+1}{\theta} \binom{n-k-1}{m-\theta} \\
&\leq \frac{1}{(1-r_1)\theta!} \times \prod_{i=0}^{\theta-1} (k+1-i) \times \prod_{i=0}^{\theta-1} (m-i) \times \frac{1}{\prod_{i=0}^{\theta-1} (n-i)} \\
&\leq \frac{1}{(1-r_1)\theta!} \times \prod_{i=0}^{\theta-1} (\frac{\sqrt{n}}{\gamma} - i) \times \prod_{i=0}^{\theta-1} (\gamma\sqrt{n} - i) \times \frac{1}{\prod_{i=0}^{\theta-1} (n-i)} \\
&\leq \frac{1}{(1-r_1)\theta!} \times \prod_{i=0}^{\theta-1} \frac{n - (\frac{1}{\gamma} + \gamma)\sqrt{n}i + i^2}{n-i} \\
&\leq \frac{1}{(1-r_1)\theta!}
\end{aligned}$$

Thus,

$$P(p_1(w) \in \mathcal{M}_1(\mathcal{D}, \theta, \gamma, T_0)) \leq \frac{1}{(1-r_1)\theta!} \leq \frac{\theta-2}{(\theta-3)\theta!}$$

We could get the same upper bound for $P(p_{i-1}(w) \in \mathcal{M}(\mathcal{D}, \theta, \gamma) | p_{i-2}(w) \in \mathcal{M}(\mathcal{D}, \theta, \gamma))$ when there are $k+1$ users containing prefix $p_i(w)$, because it is also a one step voting process to determine if $p_i(w)$ will grow on the trie.

□

### B.1 Proof of Theorem 1

*Proof.* By definition, algorithm $\mathcal{M}$ has $(\varepsilon, \delta)$-differential privacy means that, $\forall \mathcal{S} \subseteq \text{Range}(\mathcal{M})$,

$$P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}) \leq e^{\varepsilon} P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}) + \delta \tag{7}$$

When we choose the same fixed $\theta$ and $\gamma$ for a certain $n$, we abuse the notation to use $\mathcal{M}(\mathcal{D})$ and $\mathcal{M}_i(\mathcal{D}, T)$ instead of $\mathcal{M}(\mathcal{D}, \theta, \gamma)$ and $\mathcal{M}_i(\mathcal{D}, \theta, \gamma, T)$.

Suppose $|w| = l$. We decompose $\mathcal{S}$ into $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \ldots \mathcal{S}_l$. $\mathcal{S}_0$ is the subset of $\mathcal{S}$ that contains no prefix of $w$, $\mathcal{S}_1$ is the subset of $\mathcal{S}$ that contains only $p_1(w)$, $\mathcal{S}_i$ is the subset of $\mathcal{S}$ that only $p_1(w)$ to $p_i(w)$ of $w$. Formally, $\mathcal{S}_0 = \{T \in \mathcal{S} | p_i(w) \notin T, \text{ for } i = 1, 2, \ldots, l\}$ and $\mathcal{S}_i = \{T \in \mathcal{S} | p_1(w), \ldots, p_i(w) \in T \text{ and } p_{i+1}, \ldots, p_l \notin T\}$ for $i = 1, 2, \ldots, l$. Then Inequality 7 is equivalent to,

$$\sum_{i=0}^{l} P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}_i) \leq e^{\varepsilon} \sum_{i=0}^{l} P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}_i) + \delta$$

Because the tries in $\mathcal{S}_0$ do not have any node in the path of $w = (root, c1, c2, \ldots, c_l)$,

$$P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}_0) \leq P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}_0)$$

We define $\mathcal{R}_i = \{T \in \mathcal{R} | p_i(w) \in T\}$. Note that different from $\mathcal{S}_i$, $\mathcal{R}_i$ contains all possible tries that contain $p_i(w)$ (including those contain $p_{i+1}, p_{i+2}, \ldots$). Thus, $\mathcal{R}_l \subseteq \mathcal{R}_{l-1} \subseteq \cdots \subseteq \mathcal{R}_2 \subseteq \mathcal{R}_1$, therefore $P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_1) \geq P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_2) \geq \cdots \geq P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_l)$. Let $j$ be the smallest index that $P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_j) \leq \frac{\theta-2}{(\theta-3)\theta!}$, if such $j$ exists. Then,

$$\sum_{i=j}^{l} P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}_i) \leq P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_j) \leq \frac{\theta-2}{(\theta-3)\theta!}$$

For indexes $i < j$, we know that $P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_i) > \frac{\theta-2}{(\theta-3)\theta!}$. For any $i < j$:

$$P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_i) = P(p_i(w) \in \mathcal{M}(\mathcal{D})) = P(p_1(w) \in \mathcal{M}(\mathcal{D})) \times \prod_{j=1}^{i-1} P(p_{j+1} \in \mathcal{M}(\mathcal{D}) | p_j \in \mathcal{M}(\mathcal{D}))$$

Because $P(\mathcal{M}(\mathcal{D}) \in \mathcal{R}_i) > \frac{\theta-2}{(\theta-3)\theta!}$, it must be the case that every term on the right hand side of the equation above is greater than $\frac{\theta-2}{(\theta-3)\theta!}$, i.e., $P(p_1(w) \in \mathcal{M}(\mathcal{D})) \geq \frac{\theta-2}{(\theta-3)\theta!}$, and $P(p_{i-1}(w) \in \mathcal{M}(\mathcal{D}) | p_{i-2}(w) \in \mathcal{M}(\mathcal{D})) \geq \frac{\theta-2}{(\theta-3)\theta!}$, $\forall i \in \{2, \ldots, j-1\}$. By Lemma 2, $k_i > \frac{\sqrt{n}}{\gamma} - 2$, because $k_i$ is integer, $k_i \geq \frac{\sqrt{n}}{\gamma} - 1 \geq \theta$ (because $\gamma \leq \frac{\sqrt{n}}{\theta+1}$). $\forall i \in \{1, \ldots, j-1\}$.

For each $T \in \mathcal{S}_i$, $i \in \{1, \ldots, j-1\}$,

$$P(\mathcal{M}(\mathcal{D}') = T) = \prod_{b=1}^{i} P(\mathcal{M}_b(\mathcal{D}', T_{b-1}) = T_b)$$

Because $k_i > \frac{\sqrt{n}}{\gamma} - 2 \geq \theta$ for all $i \in \{1, \ldots, j-1\}$, by Lemma 1,

$$\begin{aligned}
&P(\mathcal{M}(\mathcal{D}) = T) \\
&= \prod_{b=1}^{i} P(\mathcal{M}_b(\mathcal{D}, T_{b-1}) = T_b) \\
&\leq \prod_{b=1}^{i} (1 + \frac{\theta}{k_b - \theta + 1}) P(\mathcal{M}_b(\mathcal{D}', T_{b-1}) = T_b) \\
&\leq (1 + \frac{\theta}{\frac{\sqrt{n}}{\gamma} - 1 - \theta + 1})^i P(\mathcal{M}(\mathcal{D}') = T) \\
&= (1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1})^i P(\mathcal{M}(\mathcal{D}') = T)
\end{aligned}$$

Sum up for all $T \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_{j-1}$,

$$\begin{aligned}
&P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_{j-1}) \\
&\leq (1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1})^{j-1} P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_{j-1})
\end{aligned}$$

$$\begin{aligned}
&P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}) \\
&= \sum_{i=1}^{j-1} P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}_i) + \sum_{i=j}^{l} P(\mathcal{M}(\mathcal{D}) \in \mathcal{S}_i) \\
&\leq (1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1})^{j-1} \sum_{i=1}^{j-1} P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}_i) + \frac{\theta - 2}{(\theta - 3)\theta!} \\
&\leq (1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1})^{l} \sum_{i=1}^{l} P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}_i) + \frac{\theta - 2}{(\theta - 3)\theta!} \\
&= (1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1})^{l} P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}) + \frac{\theta - 2}{(\theta - 3)\theta!} \\
&= (1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1})^{L} P(\mathcal{M}(\mathcal{D}') \in \mathcal{S}) + \frac{\theta - 2}{(\theta - 3)\theta!}
\end{aligned}$$

□

### B.2 Proof of Theorem 2

*Proof.* Suppose $\mathcal{D}$ and $\mathcal{D}'$ are user-level neighboring datasets. Without loss of generality, assume $\mathcal{D} = \{D_1, D_2, \ldots, D_n\}$ and $\mathcal{D}' = \{D_1', D_2, \ldots, D_n\}$. Let $\tilde{M}$ denote the process to first randomly selects 1 record per user (deterministically or randomly) and then applies $M$ on the sampled dataset of size $n$.

Because $M$ satisfies $(\varepsilon, \delta)$ record level DP, we know that for any record level neighboring datasets $d$ and $d'$, and $\forall \mathcal{S} \subseteq \text{Range}(M)$,

$$P(M(d) \in \mathcal{S}) \leq e^{\varepsilon} \times P(M(d') \in \mathcal{S}) + \delta$$

For any record level neighboring datasets $d$ and $d'$, without loss of generality, we will foucs on neighboring datasets that differ in the first record: $d = d_1 d_2 \ldots d_n$ and $d' = d_1' d_2 \ldots d_n$.

Our goal is to prove that $\forall \mathcal{S} \subseteq \text{Range}(\tilde{M})$, $P(\tilde{M}(D) \in \mathcal{S}) \leq e^{\varepsilon} \times P(\tilde{M}(D') \in \mathcal{S}) + \delta$. Denote $P(d|D)$ as the probability of sampling $d$ from $D$, then we can write $P(\tilde{M}(D) \in \mathcal{S})$ as:

$$\begin{aligned}
P(\tilde{M}(D) \in \mathcal{S}) &= \sum_{d} P(M(d) \in \mathcal{S}) \times P(d|D) \\
&= \sum_{d_1 d_2 \ldots d_n} P(M(d_1 d_2 \ldots d_n) \in \mathcal{S}) \times P(d_1 d_2 \ldots d_n | D) \\
&= \sum_{d_1} \sum_{d_2 \ldots d_n} P(M(d_1 d_2 \ldots d_n) \in \mathcal{S}) \times P(d_1 | D_1) \times P(d_2 \ldots d_n | D_2 \ldots D_n) \\
&= \sum_{d_2 \ldots d_n} [\sum_{d_1} P(M(d_1 d_2 \ldots d_n) \in \mathcal{S}) \times P(d_1 | D_1)] \times P(d_2 \ldots d_n | D_2 \ldots D_n)
\end{aligned}$$

Now we bound $\sum_{d_1} P(M(d_1 d_2 \dots d_n) \in \mathcal{S}) \times P(d_1|D_1)$, and then finish the proof. For any $d_1$ and $d_1'$, we know that $P(M(d_1 d_2 \dots d_n) \in \mathcal{S}) \le e^{\varepsilon} \times P(M(d_1' d_2 \dots d_n) \in \mathcal{S}) + \delta$. Thus, for a fixed $d_1$ and arbitrary $d_1'$,

$$\begin{aligned}
&\sum_{d_1} P(M(d_1 d_2 \dots d_n) \in \mathcal{S}) \times P(d_1|D_1) \\
&\le \sum_{d_1} (e^{\varepsilon} \times P(M(d_1' d_2 \dots d_n) \in \mathcal{S}) + \delta) \times P(d_1|D_1) \\
&= e^{\varepsilon} \times P(M(d_1' d_2 \dots d_n) \in \mathcal{S}) + \delta
\end{aligned}$$

Multiply both sides by $P(d_1'|D_1')$, and then sum over all $d_1'$,

$$\sum_{d_1} P(M(d_1 d_2 \dots d_n) \in \mathcal{S}) \times P(d_1|D_1) \le e^{\varepsilon} \times (\sum_{d_1'} P(M(d_1' d_2 \dots d_n) \in \mathcal{S}) \times P(d_1'|D_1')) + \delta$$

Now we finish the proof using the inequality above,

$$\begin{aligned}
P(\tilde{M}(D) \in \mathcal{S}) &= \sum_{d_2 \dots d_n} [\sum_{d_1} P(M(d_1 d_2 \dots d_n) \in \mathcal{S}) \times P(d_1|D_1)] \times P(d_2 \dots d_n | D_2 \dots D_n) \\
&\le \sum_{d_2 \dots d_n} (e^{\varepsilon} \times (\sum_{d_1'} P(M(d_1' d_2 \dots d_n) \in \mathcal{S}) \times P(d_1'|D_1')) + \delta) \times P(d_2 \dots d_n | D_2 \dots D_n) \\
&\le e^{\varepsilon} \sum_{d_2 \dots d_n} \sum_{d_1'} P(M(d_1' d_2 \dots d_n) \in \mathcal{S}) \times P(d_1'|D_1') \times P(d_2 \dots d_n | D_2 \dots D_n) + \delta \\
&= e^{\varepsilon} \times P(\tilde{M}(D') \in \mathcal{S}) + \delta
\end{aligned}$$

□

## C Proof for Corollaries

### C.1 Proof for Corollary 1

*Proof.* First get $\theta$ by standard calculation: $\theta = \max\{10, \lceil e^{W(C)+1} - \frac{1}{2} \rceil\}$, where $W$ is the Lambert $W$ function (Corless et al., 1996) and $C = (\ln \frac{8}{7\sqrt{2\pi}\delta})/e$. Then solve $L \ln(1 + \frac{1}{\frac{\sqrt{n}}{\gamma\theta} - 1}) \le \varepsilon$, we get $\gamma \le \frac{e^{\frac{\varepsilon}{L}} - 1}{\theta e^{\frac{\varepsilon}{L}}} \sqrt{n}$. Theorem 1 requires $\gamma \le \frac{\sqrt{n}}{\theta+1}$, this is satisfied by $\varepsilon \le L \ln(\theta + 1)$.

When $n \ge 10^4$, choose $\theta = \lceil \log_{10} n + 6 \rceil$. When $n = 10^4$, $\theta = 10$, and if $n$ is greater than $10^4$, it is easy to see that $\theta!$ increase faster than $n$. Formally, when $n$ increase by 10 times, $\theta$ increase by 1, and $\theta!$ increase by more than 10 times. Thus, for $n \ge 10^4$,

$$\theta! \ge \frac{n}{10^4} * 10! = \frac{n}{10^4} * 3.6 * 10^6 = 360n$$

Also when $\theta \ge 10$, $\frac{\theta-2}{\theta-3} \le \frac{8}{7}$, then,

$$\frac{\theta - 2}{(\theta - 3)\theta!} \le \frac{1}{300n}$$

□

# D Implementation of SFP

In this section we provide a description of the full algorithm of SFP in (Apple, 2017) for completeness and give the parameters of our implementation (we use the default parameters in (Apple, 2017) when provided in the paper).

SFP is based on Count Mean Sketch (CMS), which contains both client-side and server-side computations. On the client side, a string is mapped to a domain of size $m$ by one of $k$ three-wise independent hash functions. Then the client submit the result with random noise (depends on $\varepsilon$ to achieve $\varepsilon$ local DP) to the server. The server gathers results from all the clients and compute the heavy hitters. In our implementation, $m = 1024$ and $k = 2048$.

First we introduce the basis client side encoding algorithm $\mathcal{A}_{\text{client-CMS}}$. On the client side, first sample $j$ uniformly at random from $[k]$. Then construct a vector $v$ of length $m$, with $v_{h_j(d)} = -1$ and other elements in $v$ are all 1. After that, sample vector $b \in \{-1, +1\}^m$, where $b_l$ is i.i.d. and $Pr[b_l = +1] = \frac{e^{\varepsilon/2}}{e^{\varepsilon/2}+1}$. Finally, the client returns $\tilde{v} = (v_1 b_1, \ldots, v_m b_m)$ and index $j$. After receiving all the noisy hashed values from the clients, the server construct a sketch matrix, and for each element $d$, we can estimate the frequency of $d$ by a frequency oracle using the sketch matrix. This server-side algorithm is denoted as $\mathcal{A}_{\text{server-CMS}}$.

We consider strings of length up to 10 (as default in the paper) by padding shorter strings with \$ and truncating longer strings to length 10. The full SFP algorithm also contains client side and server side algorithms. There are $k = 2048$ three-wise independent hash functions with domain size of $m = 1024$, and a hash function $h$ with domain size 256 shared by the server and clients. Also, there is a threshold parameter $T$ (we used $T = 20$ and $T = 80$ in our experiments). Then apply algorithm $\mathcal{A}_{\text{client-SFP}}$ on each client, send all the results to the server and apply $\mathcal{A}_{\text{server-SFP}}$ on the server side for the final result.

In the client side algorithm $\mathcal{A}_{\text{client-SFP}}$, suppose a client holds string $s$. First sample $l$ uniformly at random from $\{1, 3, 5, 7, 9\}$, then set $r = h(s)||s[l : l + 1]$. Finally, return $\mathcal{A}_{\text{server-CMS}}(r)$ and $l$ to the server.

In the server side algorithm $\mathcal{A}_{\text{server-SFP}}$, for each $l \in \{1, 3, 5, 7, 9\}$, create sketch matrix $M_l$ by the results from the set of users submitting index $l$ and construct the frequency oracle $\tilde{f}_l$ accordingly. Also for each $l \in \{1, 3, 5, 7, 9\}$, calculate $Q_l$, which is the $T$ tuples with the largest counts $\tilde{f}_l(w||s)$ for $s \in \Omega^2$ where $\Omega$ is the 26 lowercase English letters and $w \in [256]$. For each $w \in [256]$, we form the Cartesian product of terms in $Q(w) = \{q_1|| \ldots ||q_9 : w||q_l \in Q_l$ for $l \in \{1, 3, 5, 7, 9\}$. Finally return the union of all $Q(w)$ as the result heavy hitters.

# E Additional Discussions

## E.1 Time, Space and Communication Complexity Analysis of TrieHH

**Time Complexity** Running time on the server side is $O(m)$ for each round, so the total running time is $O(mL)$. For the running time on the user side, suppose each user has at most $Z$ words, then searching for a certain prefix cost $O(ZlogZ)$. Because there are at most $\frac{m}{\theta}$ node in each level, searching for all the prefixes in this round cost at most $O(\frac{1}{\theta}mZlogZ)$. Thus the total running time for each user is $O(\frac{mZL}{\theta}logZ)$.

**Space Complexity** Space complexity on both the server and user side is the size of the trie. Because there are at most $\frac{m}{\theta}$ node in each level, and there are at most $L$ levels except the root node, total space complexity is $O(L\frac{m}{\theta})$.

**Communication Cost** The worst case communication cost for round $i$: $\frac{m^2}{\theta} \times C \times i$, where $C$ is the cost to communicate a node in the trie. This is because there are at most $\frac{m}{\theta}$ length $i$ paths in the trie at round i, and the server need to update the current trie with $m$ users. The algorithm runs for at most $L$ rounds, thus the total communication cost is at most $\sum_i \frac{m^2}{\theta} \times C \times i = \frac{m^2 L(L+1)C}{2\theta}$.

### E.2 Heavy Hitters Lists

We provide the list of the top 100 heavy hitters in the Sentiment140 dataset, and in the OOV dataset after we filter out the words in the dictionary.

**Top 100 heavy hitters with frequencies in the Sentiment140 dataset**

```
{'the': '0.1028', 'you': '0.0360', 'and': '0.0308', 'just': '0.0209', "i'm":
    '0.0169', 'for': '0.0143', 'have': '0.0107', 'going': '0.0086', 'not':
    '0.0074', 'that': '0.0073', 'was': '0.0069', 'good': '0.0062', 'work':
    '0.0056', "it's": '0.0055', 'this': '0.0053', 'watching': '0.0052', 'back':
    '0.0051', 'got': '0.0049', 'with': '0.0048', 'had': '0.0048', 'love':
    '0.0047', 'really': '0.0047', "can't": '0.0046', 'has': '0.0045', 'but':
    '0.0043', 'miss': '0.0039', 'still': '0.0039', 'its': '0.0037', 'want':
    '0.0036', 'getting': '0.0035', 'day': '0.0035', "don't": '0.0033', 'happy':
    '0.0033', 'what': '0.0032', 'now': '0.0032', 'why': '0.0031', 'lol':
    '0.0031', 'home': '0.0031', 'wish': '0.0030', 'today': '0.0030', 'all':
    '0.0029', 'new': '0.0029', 'off': '0.0028', 'need': '0.0028', 'your':
    '0.0028', 'hate': '0.0026', 'sad': '0.0026', 'last': '0.0026', 'think':
    '0.0025', 'trying': '0.0025', 'out': '0.0025', 'get': '0.0025', 'hey':
    '0.0024', 'working': '0.0023', 'like': '0.0023', 'finally': '0.0022', 'too':
    '0.0022', 'well': '0.0022', 'about': '0.0022', 'one': '0.0021', 'will':
    '0.0021', 'thanks': '0.0021', 'very': '0.0021', 'are': '0.0021', 'feel':
    '0.0020', 'cant': '0.0020', 'time': '0.0020', 'bored': '0.0020', 'feeling':
    '0.0019', 'omg': '0.0019', 'having': '0.0018', 'tired': '0.0018', 'her':
    '0.0018', 'ugh': '0.0018', 'more': '0.0017', 'waiting': '0.0017', 'missing':
    '0.0016', 'sitting': '0.0016', 'twitter': '0.0016', 'haha': '0.0016', '
    listening': '0.0016', 'how': '0.0016', 'wants': '0.0016', 'great': '0.0015',
    'wow': '0.0015', 'sick': '0.0014', 'they': '0.0014', 'know': '0.0014', 'can':
     '0.0014', 'night': '0.0014', 'another': '0.0014', 'morning': '0.0014', 'damn
    ': '0.0014', '@mileycyrus': '0.0014', 'way': '0.0014', 'yay': '0.0014', 'dont
    ': '0.0014', 'looking': '0.0013', 'some': '0.0013', 'she': '0.0013'}
```

**Top 100 heavy hitters with frequencies in the OOV dataset generated from Sentiment140**

```
{'dont': '0.011741', 'thats': '0.006008', 'didnt': '0.004292', 'sooo':
    '0.004023', 'awww': '0.003468', '@mileycyrus': '0.002931', '@tommcfly':
    '0.002556', 'soooo': '0.002473', '@ddlovato': '0.002254', 'doesnt':
    '0.001800', '#followfriday': '0.001694', 'havent': '0.001559', '
    @jonasbrothers': '0.001553', 'isnt': '0.001336', '#fb': '0.001168', 'sooooo':
     '0.001041', 'awwww': '0.001037', 'tweetdeck': '0.000958', 'couldnt':
    '0.000939', ":'(": '0.000931', 'wasnt': '0.000913', '(via': '0.000896', '
    @davidarchie': '0.000892', '@donniewahlberg': '0.000865', '@jonathanrknight':
     '0.000825', '*sigh*': '0.000811', '@jordanknight': '0.000749', 'oooh':
    '0.000730', '@mitchelmusso': '0.000708', '(and': '0.000705', 'ohhh':
    '0.000693', 'ahhhh': '0.000664', '*hugs*': '0.000647', 'nooo': '0.000634', '#
    ff': '0.000628', '#squarespace': '0.000612', 'youre': '0.000609', 'p.s':
    '0.000594', 'noooo': '0.000588', 'b/c': '0.000581', 'ughh': '0.000575', '
    goodmorning': '0.000555', 'mmmm': '0.000553', 're:': '0.000552', 'twitpic':
    '0.000540', 'sooooooo': '0.000529', '@dougiemcfly': '0.000525', '@selenagomez
    ': '0.000524', 'bgt': '0.000514', 'realised': '0.000508', "'em": '0.000503',
    'thankyou': '0.000487', "ya'll": '0.000477', 'xxxx': '0.000471', 'booo':
    '0.000464', 'youu': '0.000458', '@dannymcfly': '0.000455', 'wouldnt':
    '0.000447', 'atleast': '0.000434', 'heyy': '0.000432', "'cause": '0.000432',
    'ughhh': '0.000430', 'photo:': '0.000427', 'r.i.p': '0.000421', 'wooo':
    '0.000415', '@peterfacinelli': '0.000415', '@aplusk': '0.000409', 'tooo':
    '0.000408', 'tommorow': '0.000405', 'hayfever': '0.000405', 'a.m':
    '0.000401', '@joeymcintyre': '0.000399', 'goood': '0.000389', 'urgh':
```

```
'0.000376', '@youngq': '0.000369', 'w/o': '0.000368', 'awsome': '0.000360',
'(or': '0.000355', 'aswell': '0.000354', 'skool': '0.000354', 'tweetie':
'0.000353', 'tomorow': '0.000346', 'boooo': '0.000336', '@shaundiviney':
'0.000335', '#iranelection': '0.000335', ':-d': '0.000330', 'awwwww':
'0.000330', '#seb-day': '0.000329', 'nooooo': '0.000327', 'yeahh':
'0.000326', '@perezhilton': '0.000322', '@tomfelton': '0.000316', "g'night":
'0.000313', 'twitterverse': '0.000311', '(y)': '0.000304', 'grrrr':
'0.000299', '@officialtila': '0.000296', 'realise': '0.000289', '(not':
'0.000286', '@kirstiealley': '0.000285'}
```