New Algorithms for Text Fingerprinting

Independent University of Moscow, Moscow, Russia

CNRS, Porcet Laboratory,
Mathieu Rachout

Moscow State University, Moscow, Russia

Liaison French-Russian Institute,
Roman Kulpakov
\{ (F_\ell, i) \cdot \mathcal{C} = \mathcal{C} \mid \exists \supseteq \mathcal{C} \} = (s)F

\text{(location of } \mathcal{C} \text{ in } P) (s) \cdot s_1 \cdot \cdots \cdot s_n \cdot (s) \mathcal{C} = (s) \cdot s_1 \cdot \cdots \cdot s_n \cdot (s) \mathcal{C}

\exists \supseteq \{ (F_\ell, i) \mid \exists \supseteq \{ a \} \} = (s) \mathcal{C}

\text{Def:}

\exists \supseteq \{ u \cdot s_1 \cdot \cdots \cdot s_n = s \}

\{ b, a, \cdots, a \} = \exists

\text{Fingerprint}
Prop. \[ \exists u \geq |(s)\mathcal{J}| \geq |(s)\mathcal{F}| \]

Example: maximal location of interpreter: 
\[ \{p, q, r\} \]

Relation:\[ \mathcal{C} \not\geq \exists \frac{u_i}{s_i+1} \]

 effet: \[ \mathcal{C} \not\geq \exists \frac{s_i}{s_i-1} \]

Def. \[ \exists \mathcal{C} \subseteq \mathcal{C} \subseteq \mathcal{C} \]

Maximal Locations
\[(\|z\| \cdot \|m\|) = (z + y) - z = (m)\mathcal{J}

(1 - \alpha z) \cdot y = \|z\| \cdot \|m\|
\]

\[\forall \gamma \in \frac{\mathbb{R}}{\mathbb{Z}} \quad 1 - \gamma \alpha \frac{1}{\alpha} \gamma \alpha \frac{1}{\alpha} \gamma = \frac{1}{\alpha} \gamma \alpha \frac{1}{\alpha} \gamma \alpha \frac{1}{\alpha} \gamma = \|z\|

\cdots \gamma = \|z\| \quad \{\gamma \alpha \frac{1}{\alpha} \gamma \alpha \frac{1}{\alpha} \gamma \} = \|z\|

\exist\ n : |z| < n\] can be asymptotically less than \(|(s)\mathcal{J}|\)
Our problems

1. Compute the set \( \mathcal{F}(s) \)

2. For a given \( \mathcal{C} \subseteq \mathcal{Z} \), find all fixed points \( \mathcal{C} \)

3. In \( s \)

For a given \( \mathcal{C} \subseteq \mathcal{Z} \), find all maximal locations of \( \mathcal{C} \)
Problem 1 can be solved in $O(\min\{u, |\mathcal{X}| \log |\mathcal{X}| u\})$ time.

Problem 2 and 3 can be solved in $O(\log |\mathcal{X}|)$ time.

Problem 1 can be solved in $O(\log u)$ time.

Amir, Apostoiloa, Landau, Satta 2003:

Didier, Schmidt, Stove, Tsur 2006:
Problem 1 can be solved in time $O(\log (|s| + u))$.

Problem 2 and 3 can be solved in time and space $O(|X| + Y + |Z|)$.

Our results
<table>
<thead>
<tr>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

**Ex:** \(1010100\)

**Naming technique**

\([\_7] \leftarrow\) 

ignoreprint arrays ignoreprint names
Edges are labeled by tuples \( \{ (u, u') \} \).

\[ \begin{array}{c}
1+0u & \cdots & 1-1+0u & \cdots & 1+0u \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\end{array} \]

\[ (s) \not\in \mathcal{F} \] of ordered fingerprint names

\[ f + 0u, \ldots, f + 0u, f + I + 0u \]

Fingerprint tree
Prop: Fingerprint tree can be constructed in space \((\|s\|_F||O\) time and \(O(\log |s|_F)\) time.

Segment of length \(l + 1\) in fingerprint array

Tuple \((t, u, \overrightarrow{u}, \overrightarrow{\nu})\) points to the corresponding
(Y + \lvert \z \rvert)O$ time can be done in $O(|\z|)$ time.

Search of a given fingerprint array in fingerprint tree

$O((n + 1) \log \log n)$ time in $\{u, l, l, n\}$

The corresponding segment can be computed from
Conclusions

• \((s)^F\) on-line computation of \(F\) for \((s)^F\) time bounded for Problem 1.

• Computing integer inputs (common integer inputs)

• Sets of strings (in particular, regular languages)