The Swap Common Superstring Problem

Anna Gorbenko

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
gorbenko.ann@gmail.com

Vladimir Popov

Department of Intelligent Systems and Robotics
Ural Federal University
620083 Ekaterinburg, Russia
Vladimir.Popov@usu.ru

Abstract

In this paper we consider an approach to solve the swap common superstring problem. This approach is based on an explicit reduction from the problem to the satisfiability problem.

Keywords: swap common superstring, NP-complete, satisfiability

The algorithmic aspects of different problems of finding regularities are thoroughly studied in theoretical computer science (see e.g. [1] – [15]). In particular, the swap common superstring problem was proposed in [16].

Let
\[ \Sigma = \{a_1, \ldots, a_m\} \]
be a finite alphabet. Let
\[ S = \{S_1, \ldots, S_n\} \]
be a collection of strings over \( \Sigma \). For simplicity, we use \( S[i] \) to denote the \( i \)th letter in string \( S \), and \( S[i, j] \) to denote the substring of \( S \) consisting of the \( i \)th letter through the \( j \)th letter. The length of a string \( S \) is the number of letters in it and is denoted as \( |S| \). Let
\[ \#occ(X, Y) = |\{i \mid X = Y[i, j]\}|. \]
The decision version of the swap common superstring problem can be formulated as following.

**THE SWAP COMMON SUPERSTRING PROBLEM (SWCS):**

**Instance:** A collection $S$ of strings over $\Sigma$, a string $T$, and a positive integer $k$.

**Question:** Is there a string $S$ such that

$$|S| = |T|,$$

$$\{S[i], S[i+1]\} = \{T[i], T[i+1]\},$$

for all $1 \leq i < |S|$, and

$$|\{i \mid \#occ(S_i, S) \geq 1\}| \geq k?$$

The problem SWCS is \textbf{NP}-hard [16]. Encoding different hard problems as instances of SAT and solving them with efficient satisfiability algorithms has caused considerable interest (see e.g. [17] – [36]). In this paper, we consider an approach to solve the SWCS problem. Our approach is based on an explicit reduction from the problem to the satisfiability problem.

Let

$$\varphi[1] = \land_{1 \leq i \leq |T|} \lor_{1 \leq j \leq m} x[i, j],$$

$$\varphi[2] = \land_{1 \leq i \leq |T|} \land_{1 \leq j[1], j[2] \leq m} \neg x[i, j[1]] \lor \neg x[i, j[2]],$$

$$\varphi[3] = \land_{1 \leq i \leq |T|, 1 \leq j \leq m, T[i] \neq a_j} ((\neg x[i, j] \lor x[i+1, s]) \land (\neg x[i, s] \lor x[i+1, j]),$$

$$\varphi[4] = \land_{1 \leq i \leq k} \land_{1 \leq j \leq n} y[i, j],$$

$$\varphi[5] = \land_{1 \leq i \leq k, 1 \leq j[1], j[2] \leq n} (\neg y[i, j[1]] \lor \neg y[i, j[2]],$$

$$\varphi[6] = \land_{1 \leq i \leq k, 1 \leq j \leq n} \land_{1 \leq p \leq |T|-|S_j|+1} (\neg y[i, j] \lor \neg y[i, j] \lor \neg x[j, p])],$$

$$\varphi[7] = \land_{1 \leq i \leq k, 1 \leq j \leq n} \land_{1 \leq p \leq |T|-|S_j|+1} (\neg y[i, j] \lor \neg z[j, p[1]] \lor \neg z[j, p[2]],$$

$$\varphi[8] = \land_{1 \leq i \leq k, 1 \leq j \leq n} \land_{1 \leq p \leq |T|-|S_j|+1, p \leq q \leq p+|S_j|-1, 1 \leq r \leq m, S_j[q-p+1] \neq a_r \neq a_r} (\neg y[i, j] \lor \neg x[q, r].$$
The swap common superstring problem

\[ \xi = \land_{i=1}^{8} \varphi[i]. \]

It is easy to check that there is a string \( S \) such that

\[ |S| = |T|, \]

\[ \{S[i], S[i+1]\} = \{T[i], T[i+1]\}, \]

for all \( 1 \leq i < |S| \), and

\[ |\{i \mid \#occ(S_i, S) \geq 1\}| \geq k \]

if and only if \( \xi \) is satisfiable. It is clear that \( \xi \) is a CNF. So, \( \xi \) gives us an explicit reduction from SWCS to SAT. Now, using standard transformations (see e.g. [37]) we can obtain an explicit transformation \( \xi \) into \( \zeta \) such that \( \xi \Leftrightarrow \zeta \) and \( \zeta \) is a 3-CNF. Clearly, \( \zeta \) gives us an explicit reduction from SWCS to 3SAT.

We have designed a generator of natural instances for SWCS. We have considered our genetic algorithms OA[1] (see [38]) and OA[2] (see [39]) for SAT. For solution of SWCS, we have used heterogeneous cluster. Each test was runned on a cluster of at least 100 nodes. Selected experimental results are given in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>time</th>
<th>average</th>
<th>max</th>
<th>best</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA[1]</td>
<td>1.22 h</td>
<td>11.41 h</td>
<td>3.17 min</td>
<td></td>
</tr>
<tr>
<td>OA[2]</td>
<td>39 min</td>
<td>6.23 h</td>
<td>6.54 min</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Experimental results for SWCS.

**ACKNOWLEDGEMENTS.** The work was partially supported by Analytical Departmental Program “Developing the scientific potential of high school” 8.1616.2011.

**References**


Received: November 1, 2012