Notes on the paper
“Short Labels by Traversal and Jumping”

Cyril Gavoille

August 4, 2011

Abstract

This is a correction to the labeling scheme for Caterpilars, after the remarks raised by Martin Böhm\(^1\). The new label length is 1 bit larger.

1 Labeling the Caterpilars

The scheme is essentially the same, mainly its analysis is changed. We use the same notations. The main difference is that the length \( p_i \) of the interval where \( \ell(x_i) \) is computed is set to \( p_i = 2t_i^+ + 4 \) instead of \( p_i = 2t_i^+ + 3 \). For \( C(x_i) \) we keep its initial definition, that is:

\[
C(x_i) = \text{code}_0(t_i + 3 - |\text{code}_1(t_{i+1})|) \circ \text{code}_1(t_{i+1})
\]

The adjacency scheme is the same. Here are the constraints that all \( t_i \) and \( p_i \) must fulfill:

1. \( t_i + 3 - |\text{code}_1(t_{i+1})| \geq 0 \) and \( t_{i+1} \geq 0 \), because of the definition of \( C(x_i) \).
2. \( p_i \geq 2^{\lvert C(x_i) \rvert} \) because \( \ell(x_i) \) must encode \( C(x_i) \) thanks to the coding Claim.
3. \( p_i \geq d_i + 1 \) because the interval \([\ell(x_i), \ell(x_i) + p_i]\) must be large enough to contain at least \( \ell(x_i) \) and the labels of its \( d_i \) leaves attached to \( x_i \).

Condition 2 is equivalent to \( \lg p_i \geq |C(x_i)| = 1 + (t_i + 3 - |\text{code}_1(t_{i+1})|) + |\text{code}_1(t_{i+1})| \), that is true since precisely we have set \( \lg p_i = t_i + 4 \). Condition 3 is equivalent to \( t_i + 4 \geq \lg (d_i + 1) \).

All these conditions are fulfilled if we set:

\[
t_i = \max \{0, |\text{code}_1(t_{i+1})| - 3, \lceil \lg d_i \rceil - 4\}, \quad \text{with } t_{k+1} = 0.
\]

Appart the extra bit added to the label to distinguish inner vertices and leaves, the maximum label assigned by the scheme above is no more than \( 2 \sum_{i=1}^{k} p_i \). The analysis of \( \sum p_i \) is slightly simpler, although quite similar to the original one.

\(^1\)bohmatrey@karlin.mff.cuni.cz
Let us proved property \((P_m)\) that is:

\[
\sum_{i=m}^{k} p_i \leq \left( 2^5 \sum_{i=m}^{k} \lceil d_i + 1 \rceil \right) - p_m.
\]

Recall that \(\lfloor x \rfloor_2 = 2^{\lceil \lg x \rceil}\). Note that by definition of \(t_i\), only three values are possible for \(p_i\):

\[
p_i = 2^{4+t_i} = 2^{4 + \max\{0, |\text{code}_1(t_{i+1})| - 3, \lceil \lg d_i \rceil - 4\}}
\]

\[
= \max\{2^4, 2^{\text{code}_1(t_{i+1})+1}, \lceil d_i + 1 \rceil \}.
\]

For \((P_k)\), the left term is \(A = p_k\) whereas the right term is \(B = 2^5 \lceil d_k + 1 \rceil - p_k\). Thus \(A \leq B\) iff \(p_k \leq 2^4 \lceil d_k + 1 \rceil\).

1. If \(p_k = \lceil d_k + 1 \rceil\), then we are done.
2. If \(p_k = 2^4\), then we are done since \(\lceil d_i + 1 \rceil \geq 1\) for all \(d_i \geq 0\).
3. If \(p_k = 2^{\text{code}_1(t_{k+1})+1}\). Note that \(t_{k+1} = 0\) and that \(|\text{code}_1(0)| = 2\). Thus \(p_k = 2^3\), and we are done.

Therefore, \((P_k)\) is proved. Assume that \((P_i)\) is true for all indices \(i \in \{m, \ldots, k\}\) for some \(m > 1\). Let us prove \((P_{m-1})\).

The left term is now \(A = \sum_{i=m-1}^{k} p_i\) which by assumption is

\[
A = p_{m-1} + 2^5 \sum_{i=m}^{k} \lceil d_i + 1 \rceil_2 - p_m.
\]

The right term is \(B = 2^5 \sum_{i=m-1}^{k} \lceil d_i + 1 \rceil_2 - p_{m-1}\) which is also

\[
B = 2^5 \lceil d_{m-1} \rceil_2 + 2^5 \sum_{i=m-1}^{k} \lceil d_i + 1 \rceil_2 - p_{m-1}.
\]

Hence, \(A \leq B\) iff \(p_{m-1} - p_m \leq 2^5 \lceil d_{m-1} \rceil_2 - p_{m-1}\) that is also:

\[
p_{m-1} \leq \frac{1}{2} p_m + 2^4 \lceil d_{m-1} + 1 \rceil_2.
\]

Let us prove the latter equation.

1. If \(p_{m-1} = \lceil d_{m-1} + 1 \rceil_2\) or \(2^4\), then we are done.
2. If \(p_{m-1} = 2^{\text{code}_1(t_m)+1}\). We observe that for every integer \(x \geq 0\), \(|\text{code}_1(x)| = 2|\text{bin}(x)|\). Moreover we check that \(2|\text{bin}(x)| \leq x + 2\). The first values for \(2|\text{bin}(x)|\) are \(2, 2, 4, 4, 6, 6, 6, 8, \ldots\) whereas for \(x + 2\) there are: \(2, 4, 6, 8, 10, \ldots\).

In particular, \(2^{\text{code}_1(t_m)+1} \leq 2^{t_{m+3}} = \frac{1}{2} p_m\). It follows that \(p_{m-1} \leq \frac{1}{2} p_m\), and we have proved that \((P_m)\) is true.
Let \( \ell(n) \) be the maximum label assigned by the scheme above. We have \( \ell(n) \leq 2 \sum_{i=1}^{k} p_i \leq 2^6 \sum_{i=1}^{k} [d_i + 1] \). We observe that for all \( x \geq 0, \lfloor x + 1 \rfloor_2 \leq \max\{2x, 1\} \).

Thus,

\[
\ell(n) \leq 2^6 \cdot \sum_{i=1}^{k} \max\{2d_i, 1\} = 2^6 \cdot \max\left\{ 2 \sum_{i=1}^{k} d_i, k \right\} = 2^6 \cdot \max\{2(n-k), k\}
\]

since \( \sum_{i=1}^{k} d_i = n - k \). If \( k \geq 2n/3 \), then \( \max\{2(n-k), k\} = k \leq n \). If \( k < 2n/3 \), then \( \max\{2(n-k), k\} = 2(n-k) < 2n/3 \). It follows that \( \ell(n) \leq 2^6 \cdot n \).

In conclusion, the label length is no more than \( \lceil \log n \rceil + 6 + 1 \) bits.