One-and-a-Half-Side Boundary Labeling

Chun-Cheng Lin[1,⋆], Sheung-Hung Poon[2,⋆⋆], Shigeo Takahashi[3], Hsiang-Yun Wu[3], and Hsu-Chun Yen[4,⋆⋆⋆]

[1] Dept. of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan
[2] Dept. of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan
[3] Dept. of Complexity Science and Engineering, The University of Tokyo, Kashiwa-city, Chiba 277-8561, Japan
[4] Dept. of Electrical Engineering, National Taiwan University, Taipei 106, Taiwan

Abstract. In *boundary labeling*, each point site in a rectangular map is connected to a label outside the map by a *leader*, which may be a rectilinear or a straight-line segment. Among various types of leaders, the so-called type-*opo* leader consists of three segments (from the site to its associated label) that are orthogonal, then parallel, and then orthogonal to the side to which the label is attached. In this paper, we investigate the so-called *1.5-side boundary labeling*, in which, in addition to being connected to the right side of the map directly, type-*opo* leaders can be routed to the left side temporarily and then finally to the right side. It turns out that allowing type-*opo* leaders to utilize the left side of a map is beneficial in the sense that it produces a better labeling result in some cases. To understand this new version of boundary labeling better, we investigate from a computational complexity viewpoint the *total leader length minimization* problem as well as the *bend minimization* problem for variants of *1.5-side boundary labeling*, which are parameterized by the underlying label size (uniform vs. nonuniform) and port type (fixed-ratio, fixed-position, vs. sliding). For the case of nonuniform labels, the above two problems are intractable in general. We are able to devise pseudo-polynomial time solutions for such intractable problems, and also identify the role played by the number of distinct labels in the overall complexity. On the other hand, if labels are identical in size, both problems become solvable in polynomial time. We also characterize the cases for which utilizing the left side for routing type-*opo* leaders does not help.

Keywords: Map labeling, boundary labeling, complexity.

1 Introduction

In map labeling [5,10,11], the basic requirement for placing labels in a map is that all the labels should be pairwise disjoint. It is clear that such a requirement is

⋆ Corresponding author (E-mail: `cclin321@nctu.edu.tw`). Research supported in part by NSC 98-2218-E-151-004-MY3.
⋆⋆ Research supported in part by NSC 97-2221-E-007-054-MY3.
⋆⋆⋆ Research supported in part by NSC 97-2221-E-002-094-MY3.

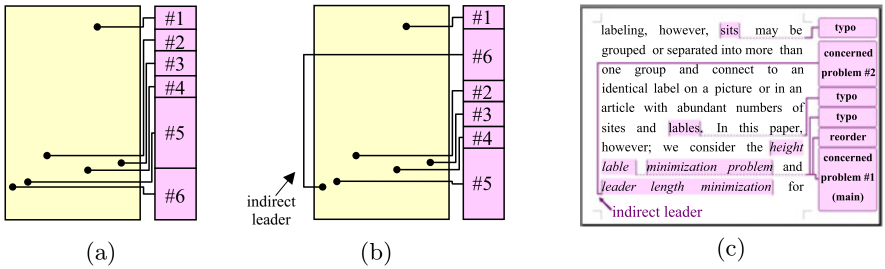


Fig. 1. (a) One-side boundary labeling with type-*opo* leaders. (b) 1.5-side boundary labeling with type-*opo* leaders. (c) Application to a word processing annotation system.

difficult to be achieved in the case where large labels are placed on dense points. To address this problem, Bekos et al. [3] proposed the so-called *boundary labeling*, in which all labels are attached to the boundary (four sides) of a rectangle R enclosing all sites, and each site is connected to a unique label by a *leader*, which may be a rectilinear or a straight-line segment. In such a setting, they assumed no two sites with the same x- or y- coordinates, and investigated how to place the labels and leaders in a drawing such that there are no crossings among leaders and either the total leader length or the total number of bends of leaders is minimized under a variety of constraints. Bekos et al. [2] investigated a similar problem for labeling polygonal sites under the framework of boundary labeling. Subsequently, Lin [7] used hyperleaders and multiple copies of the same label to propose algorithms for crossing-free multi-site-to-one-label boundary labeling [8], in which more than one site is allowed to be connected to a common label.

Boundary labeling [3,4,1] is characterized as *k-side labeling with type-t leaders* (where $k \in \{1, 2, 4\}$ and $t \in \{opo, po, s, do\}$) if the labels are allowed to attach to the k sides of the enclosing rectangle R by only type-t leaders. The parameter t specifies the way in which a leader is drawn to connect a site to a label. The *opo*, *po*, s, and *do* stand for *orthogonal-parallel-orthogonal*, *parallel-orthogonal*, *straight-line* and *diagonal-orthogonal* leader types, respectively. It is assumed that the parallel (i.e., 'p') segment associated with a type-*opo* leader lies in a *track routing area* sandwiched between R and the label stack. See Figure 1(a). In a recent article [9], the idea of the so-called *indirect* leaders was proposed for possibly shortening the total leader length in one-side boundary labeling. *Indirect* leaders utilize the left side of the map to route the parallel segments of type-*opo* leaders (while the original ones that do not utilize the left side are called *direct* leaders). See Figure 1(b) for an example. Such a new type of boundary labeling is called *1.5-side boundary labeling with type-opo leaders*, which is likely to find applications in, for example, text annotation for word processing softwares as [9] suggests (see Figure 1(c) for an illustrating example). In general, the track routing area is not wide, so too many direct leaders in this area would make it difficult to tell them apart (see Figure 1(a)). The introduction of indirect leaders may lead to fewer direct leaders in this area, where the routing of those direct leaders can be distinguished more easily (see Figure 1(b)).

In this paper, we consider, from a computational complexity viewpoint, the *total leader length minimization* (*TLLM*) problem and the *total bend minimization* (*TBM*) problem for variants of *1.5-side boundary labeling with type-opo leaders*, which are parameterized by their label and port types. A label can be of *uniform* or *nonuniform* size, and the port (i.e., the position where a leader touches a label) associated with a label is of type *fixed-ratio*, *fixed-position*, or *sliding*. It turns out that for nonuniform labels, both *TLLM* and *TBM* are intractable in general regardless of the port type. We are able to design a pseudo-polynomial time algorithm and a fixed-parameter algorithm for such intractable problems. Both *TLLM* and *TBM* become solvable in polynomial time if labels are of uniform size. Interestingly, we also show that for labels of uniform size and under either sliding or fixed-ratio port type, indirect leaders do not help as far as minimizing the total leader length is concerned.

2 Preliminaries

2.1 The Models for 1.5-Side Boundary Labeling

In 1.5-side boundary labeling, we assume that sites are points of zero size located on the plane, and only type-*opo* leaders are used, no matter whether they are direct or indirect. Following Bekos et al.'s convention [2], various models for 1.5-side boundary labeling can be differentiated according to a triple (*LabelSize*, *LabelPort*, *Objective*), where:

LabelSize: Each label l_i is associated with a height h_i and a width w_i. As each leader is connected to the left side of a label box, w.l.o.g., we assume that $\forall 1 \leq i, j \leq n, w_i = w_j$, where n is the number of labels. Labels are of *uniform* size if $\forall 1 \leq i, j \leq n, h_i = h_j$; otherwise, of *nonuniform* size.

LabelPort: Depending on the location where a leader touches a label, consider the following three types:

- *Fixed-ratio port* (*FR* for short): there exists a constant $0 \leq \alpha \leq 1$, such that the i-th leader touches the point of height αh_i, from the bottom of the i-th label. In [8], α is assumed to be $\frac{1}{2}$, i.e., each leader touches the middle of the corresponding label.
- *Fixed-position port* (*FP* for short): The i-th label is associated with a predefined constant $0 \leq \alpha_i \leq 1$ such that the i-th leader touches the point of height $\alpha_i h_i$, from the bottom of the i-th label.
- *Sliding port*: As the name suggests, the contact point of a leader can *slide* along the corresponding label edge.

Objective: Find a legal label placement such that the total leader length is minimum (*TLLM*), or the total number of bends is minimum (*TBM*).

Before proceeding further, we first show three examples as depicted in Figure 2, which suggests that indirect leaders really help.

Our main results are given in Table 1, where the uniform-label cases can be solved in polynomial time, while the nonuniform-label cases are NP-complete.

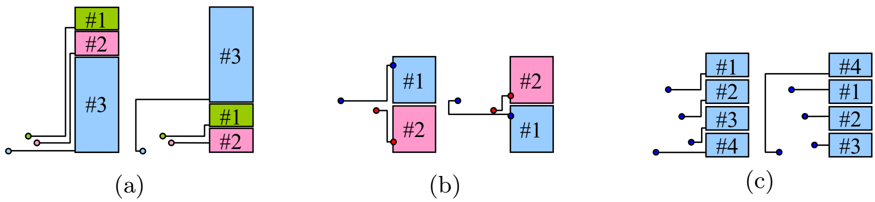


Fig. 2. (a) *TLLM* with nonuniform labels; (b) *TLLM* with uniform labels and *FP* ports; (c) *TBM* with uniform labels

Table 1. Time complexity for a variety of 1.5-side boundary labeling models. *FR* denotes fixed-ratio port, and *FP* denotes fixed-position port.

(*LabelSize*,	*LabelPort*,	*Objective*)	time	reference
(*uniform*,	*FR*/*sliding*,	*TLLM*)	$O(n \log n)$	Thm 1
(*uniform*,	*FP*,	*TLLM*)	$O(n^5)$	Thm 2
(*uniform*,	*FR*/*FP*/*sliding*,	*TBM*)	$O(n^5)$	Thm 2
(*nonuniform*,	*FR*/*FP*/*sliding*,	*TLLM*)	NP-complete*	Thm 3
(*nonuniform*,	*FR*/*FP*/*sliding*,	*TBM*)	NP-complete*	Thm 4

* An $O(n^4 h)$ pseudo-polynomial time algorithm is developed in Theorem 5, where h is the height of the map. Furthermore, if labels are of k different heights, an $O(n^{k+4})$ time algorithm is available as Theorem 6 shows.

2.2 Problem Setting

We consider the following labeling problem. Given a rectangular area R of height h and width w whose left lower corner resides at the origin of the x-y plane (i.e., $R = [0, w] \times [0, h]$), and a set of n points (called *sites*) $p_i = (x_i, y_i), 1 \leq i \leq n$, located inside R (i.e., $0 \leq x_i \leq w, 0 \leq y_i \leq h, 1 \leq i \leq n$), each of which is associated with a rectangular label l_i of width w_i and height h_i, the *one-and-a-half-side* (1.5-side, for short) *boundary labeling* problem is to place the labels along one side of the boundary of R, and connect p_i to l_i, $1 \leq i \leq n$ using *rectilinear leaders* that are either direct or indirect (or leaders, for short) so that a certain criterion is met. As illustrated in Figure 1, a *rectilinear leader* consists of horizontal and/or vertical line segments connecting a site to its corresponding label. We assume that h, w, x_i, y_i, h_i, w_i, $1 \leq i \leq n$, are all positive integers. We further require that a leader has at most two bends, which occur only in one of the two *track routing areas* denoted as A_{left} and A_{right} (see Figure 3). A leader bending in the area A_{right} (resp., A_{left}) is called a *direct* (resp., *indirect*) leader. Throughout the rest of this paper, we assume that there are no two sites with the same x- or y- coordinate, and sites are labeled as $p_1, p_2, \cdots, p_n$ in the increasing order of their y-coordinates.

We assume that $\sum_{i=1}^{n} h_i = h$, i.e., the label heights sum up to the height of R. In this case, the y-coordinate of the top of the j-th label is $\sum_{i=1}^{j} h'_i$, where h'_i is the height of the i-th label from the bottom of the label stack. Note

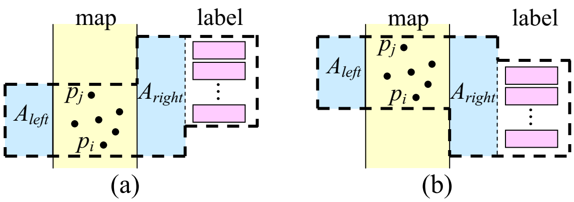


Fig. 3. Illustration of two cases for $G(i, j)$

that since we allow indirect leaders, label l_j need not be the j-th label. For the criteria considered in this paper, we may further assume that labels are of uniform width (i.e., $w_i = w_j$, $\forall 1 \leq i, j \leq n$) with no loss of generality. Consider a region $G(i, j)$ induced by a subset of sites $p_i, \cdots, p_j$ and their corresponding labels $l_i, \cdots, l_j, 1 \leq i < j \leq n$ as shown in Figure 3. Note that the left track routing area is of height less than $(y_{j+1} - y_{i-1})$, and the height of the right track routing area depends on the relative positions of the sites $p_i, \cdots, p_j$ and their labels. A leader connecting a site p_k to its corresponding label l_k (where $1 \leq k \leq j$) is *legal* with respect to $G(i, j)$ if it resides entirely in $G(i, j)$.

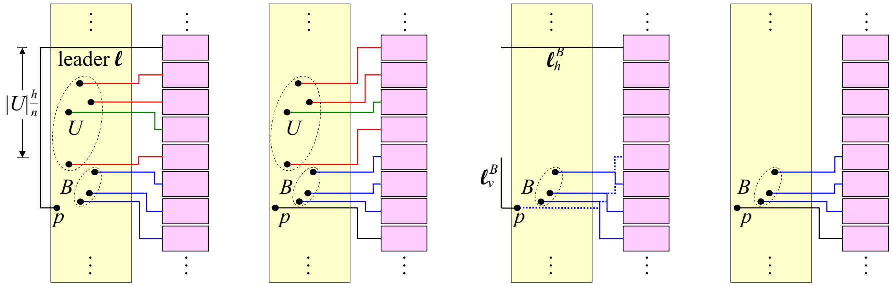
3 Uniform-Label Cases

Consider the uniform-label case, i.e., $h_i = \frac{h}{n}, \forall 1 \leq i \leq n$. As Example 2(b) indicates, using indirect leaders may result in a shorter total leader length in some cases, provided that leaders are connected to fixed-position ports. In contrast, if the ports are fixed-ratio (with respect to all labels) or can slide along boundaries of labels, the following result shows indirect leaders to be unnecessary:

Lemma 1. *In the case of uniform labels, direct leaders are sufficient to achieve optimal solutions with respect to TLLM under either the fixed-ratio port or the sliding port model.*

Proof (Sketch). In the following, we only show the fixed-ratio port model, because the proof for the sliding port model is similar. It suffices to prove that the total leader length of any labeling L_i with i indirect leaders is no shorter than that of the labeling δ with only direct leaders. Let $\phi(L)$ denote the total leader length of labeling L. The basic idea of our proof is to find labelings L_{i-1}, L_{i-2}, $\ldots$, L_1 such that $\phi(L_i) \geq \phi(L_{i-1}) \geq \cdots \geq \phi(L_1) \geq \phi(\delta)$, where L_j denotes a labeling with j indirect leaders for $1 \leq j < i$.

Consider the inner-most indirect leader (i.e., the indirect type-*opo* leader has the rightmost parallel segment) in labeling L_i, which is denoted by ℓ, and the site connected with ℓ is denoted by p (see Figure 4(a)). Hence, all the sites wrapped by ℓ are connected only by direct leaders, and they are divided into two groups according to the orientation of their associated direct leaders. Consider each of those wrapped sites from the bottom to the top, say site q. From site q to its corresponding label, if the vertical segment of the associated leader goes



(a) Labeling L_i. (b) Labeling L_{i-1}. (c) (d)

Fig. 4. The inner-most indirect leader in labeling (a) is replaced by a direct leader in labeling (b). (c) and (d) are respectively the labeling (a) and (b) where the total leader length of (b) for those sites in U are removed. Note that the leader segments that occur in (d) but not in (c) are represented by dotted-line segments in (c).

downward and all the sites located below q belong to group B, then $q \in B$; otherwise, $q \in U$ (see Figure 4(a)). That is, there exist sites with leaders going downward but not belonging to group B. Let $\Delta = \{p\} \cup U \cup B$.

If $\Delta = \{p\}$ only, then indirect leader ℓ can be changed to a direct one immediately, and the modified labeling is our required labeling L_{i-1}. Otherwise, we establish a labeling L_{i-1} with $(i-1)$ indirect leaders which is almost the same as labeling L_i except all the sites in Δ are connected by direct leaders (see Figure 4(b)). Let $d(p_j)$ (resp., $d'(p_j)$) denote the length of the leader connected to site p_j in L_i (resp., L_{i-1}), for any $j \in \{1, \cdots, n\}$. For any set of sites, let $d(A) = \sum_{p_j \in A} d(p_j)$, and define $d'(A)$ similarly. From Figures 4(a) and 4(b), it is observable that $d(U) + |U|\frac{h}{n} \geq d'(U)$. Since $d(p) > |U|\frac{h}{n}$ in labeling L_i (see Figure 4(a)), the lengths $d(U) + |U|\frac{h}{n}$ and $d'(U)$ can be removed in L_i and L_{i-1}, respectively. Figure 4(c) (resp., Figure 4(d)) is the labeling L_i in (a) (resp., L_{i-1} in (b)) removing the length $d(U)+|U|\frac{h}{n}$ (resp., $d'(U)$), i.e., the total leader length of L_i is decreased no less than that of L_{i-1}. Note that in Figure 4, the leader segments that occur in (d) but not in (c) are represented by dotted-line segments in (c). It is easy to see that the remaining length of the indirect leader ℓ in (c) can be used to compensate those dotted-line segments in (c), in which lengths ℓ_h^B and ℓ_v^B compensate dotted horizontal and vertical segments, respectively. Even so, there still exists a nonnegative remaining length of the indirect leader ℓ in (c). Hence, $\phi(L_i) \geq \phi(L_{i-1})$, as required.

Like the above, we can construct L_{j-2} according to L_{j-1} such that $\phi(L_{j-1}) \geq \phi(L_{j-2})$, and so forth. Finally, we have $\phi(L_i) \geq \phi(L_{i-1}) \geq \cdots \geq \phi(L_1) \geq \phi(\delta)$, as required. □

According to Lemma 1 above and [3] (the boundary labeling with direct leaders can be found in $O(n \log n)$ time), we have the following theorem.

Theorem 1. *The* 1.5*-side boundary labeling of the model* (*uniform*, *FR*/*sliding*, *TLLM*) *can be found in* $O(n \log n)$ *time.*

Next, we use a dynamic programming strategy to solve both *TLLM*[1] and *TBM* for 1.5-side boundary labeling with type-*opo* leaders in $O(n^5)$ time, regardless of the underlying port model. Note that according to Theorem 1, *TLLM* can be solved in $O(n \log n)$ time under the *FR/sliding* port model.

Theorem 2. *With respect to opo-type* 1.5*-side boundary labeling, both TLLM and TBM can be solved in* $O(n^5)$ *time when labels are of uniform height, regardless of the underlying port model.*

Proof. (Sketch) In what follows, we only consider *TLLM*; the *TBM* case is similar.

The proof is based on the strategy of dynamic programming. Without loss of generality, we assign indices j $(1 \leq j \leq n)$ to labels in the label stack in the increasing order of their y-coordinates. Note that due to the nature of 1.5-side boundary labeling, label l_i (i.e., the one associated with site p_i) need not be the i-th label in the label stack. It is easy to see that the y-coordinates of the top and bottom edges of the i-th label from the bottom of the label stack are $i \times \frac{h}{n}$ and $(i-1) \times \frac{h}{n}$, respectively. In the case of using fixed-ratio ports, we define $\gamma(p_i, j)$ (resp., $\ell(p_i, j)$) to be the length of a direct (resp., indirect) type-*opo* leader from site p_i to the fixed-position port of the j-th label in the label stack. In the case of using sliding ports, $\gamma(p_i, j)$ and $\ell(p_i, j)$ are defined similarly except that a leader connects p_i to the closest point (as opposed to the fixed-position point) on the boundary of the j-th label.

Let $S(a, b, c)$ (where $1 \leq a \leq b \leq n, 1 \leq c \leq n$) denote the minimal total leader length where sites $p_a, \cdots, p_b$ $(1 \leq a \leq b \leq n)$ are connected to the labels of the label stack that are consecutive in the bottom to top order starting from the c-th one up to the $(c-(b-a))$-th one using only legal leaders. We define $R_{a,b}$ to be the rectangle $[0, w] \times [y_{a-1}, y_{b+1}]$. Note that $R_{a,b}$ defines the left track routing area through which any indirect leader from site $p_i, a \leq i \leq b$ can be routed. In other words, it is illegal for p_i to use an indirect leader to connect to a label whose y-coordinate is above y_{b+1} or below y_{a-1}.

Our dynamic programming formula for $S(a, b, c)$ is as follows (see Figure 5):

$$
\begin{aligned}
\min\{&\sum_{i=0}^{b-a} \gamma(p_{a+i}, c+i), \\
&\min_{i,j \in \{0,\cdots,b-a\}, j<i} \{\ell(p_{a+i}, c+j) + S(a, a+j-1, c) \\
&\qquad + S(a+j, a+i-1, c+j+1) + S(a+i+1, b, c+i+1)\}, \\
&\min_{i,j \in \{0,\cdots,b-a\}, j>i} \{\ell(p_{a+i}, c+j) + S(a, a+i-1, c) \\
&\qquad + S(a+i+1, a+j, c+i) + S(a+j+1, b, c+j+1)\}\}
\end{aligned}
$$

In view of the above, it is reasonably easy to see that the minimum total leader length equals $S(1, n, 1)$.

[1] A dynamic programming formulation was originally given for a simpler version of *TLLM* in [9].

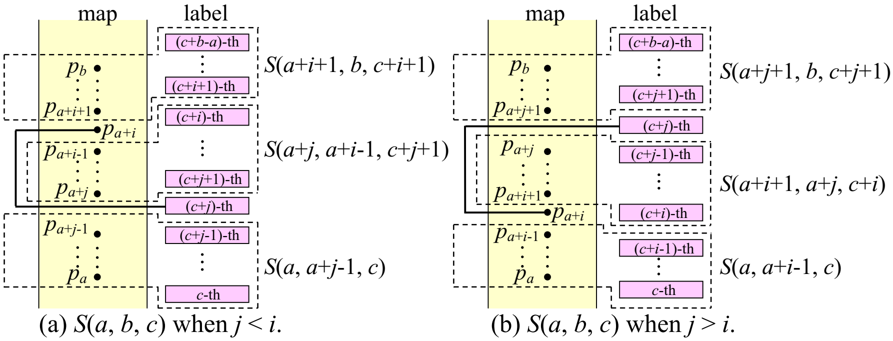


Fig. 5. Illustration of a subproblem

As for the complexity of the algorithm, we construct n tables $T_1, \cdots, T_n$ where table T_i has $(n-i+1) \times (n-i+1)$ entries for $i \in \{1, \cdots, n\}$; the entry (j,k) in table T_i records the minimal total leader length when sites $p_j, p_{j+1}, \cdots, p_{j+i-1}$ are connected to the k-th, the $(k+1)$-th, $\cdots$, and the $(k+i-1)$-th labels in the label stack by direct and indirect type-*opo* leaders. By doing so, the solution of our problem can be found in the entry $T_n(1,1)$. In view of the dynamic programming formula, each entry of Table $T_i, 1 \le i \le n$, can be computed in $O(i^2)$ time. Hence, using a bottom up approach, $T_n(1,1)$ can be obtained in time $O(\sum_{i=1}^{n}(n-i+1)^2 i^2) = O(n^5)$. □

4 Nonuniform-Label Cases

4.1 NP-Hardness

In this subsection, we show that the 1.5-side boundary labeling for non-uniform labels is NP-complete for both *TLLM* and *TBM*.

Theorem 3. *It is NP-complete to find a* 1.5*-side boundary labeling of model* (*non- uniform*, *FR*/*FP*/*sliding*, *TLLM*).

Proof (Sketch). We only consider the *FR*/*FP*-port model here. The NP-complete proof for the sliding-port model can be shown along a similar line, and therefore, is omitted here. To see that the problem is in NP, since we assume that h, w, x_i, y_i, h_i, w_i, $1 \le i \le n$, are all positive integers. In order to show the NP-hardness of our problem, we obtain a linear-time reduction from a single-machine scheduling problem, called *total discrepancy problem* [6], to our problem. On one machine, we plan to arrange the schedule for the non-preemptive execution of a set J of $2n+1$ jobs $J_0, J_1, \ldots, J_{2n}$. Each job J_i has an execution time length $l_i \in \mathbb{Z}^+$ such that $l_0 < l_1 < \ldots < l_{2n}$. For a planned schedule σ, the actual execution midtime for job J_i is denoted by $m_i(\sigma)$. Each job has a *preferred midtime*, which corresponds to the time at which we would like the first half of the job to be completed. We assume that all the jobs share a single preferred

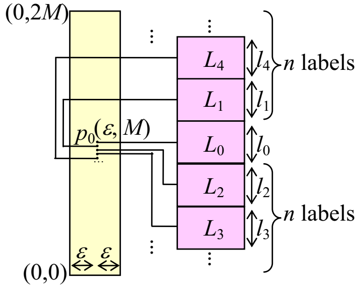


Fig. 6. Reduction in Theorem 3

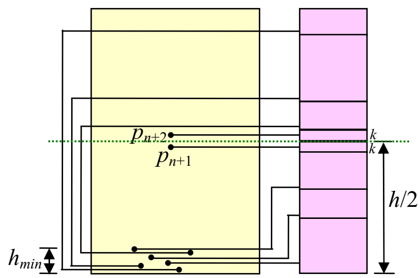


Fig. 7. Reduction in Theorem 4

midtime $M = \sum_{i=0}^{2n} l_i/2$. The *penalty* of job J_i for a schedule σ is defined as the absolute difference of its midtime to its preferred midtime, i.e., $|m_i(\sigma) - M|$ for $0 \leq i \leq 2n$. The *cost* incurred in a schedule σ is then defined to be the total penalties incurred by all jobs. The objective of the total discrepancy problem is to determine a schedule σ such that the total cost of the schedule, i.e., $\sum_{i=0}^{2n} |m_i(\sigma) - M|$, is minimized. Garey, Tarjan and Wilfong [6] showed the following properties for an optimal schedule σ_{opt} of the $2n+1$ jobs $J_0, J_1, \cdots, J_{2n}$, but the decision problem whether such a schedule exists is NP-complete.

1. σ_{opt} does not have any gaps between the jobs.
2. The midtime of the shortest job J_0 is M, i.e., $m_0(\sigma_{opt}) = M$.
3. Jobs $J_1, J_2, \ldots, J_{2n}$ are divided into two groups, $A(\sigma_{opt}) = \{J_i : m_i(\sigma_{opt}) < M\}$ and $B(\sigma_{opt}) = \{J_i : m_i(\sigma_{opt}) > M\}$, such that $|A(\sigma_{opt})| = |B(\sigma_{opt})| = n$.
4. Suppose the sequence of the jobs in schedule σ_{opt} is $A_n, A_{n-1}, \ldots, A_1, J_0, B_1, B_2, \ldots, B_n$. Then $\{A_i, B_i\} = \{J_{2i-1}, J_{2i}\}$.
5. The optimal cost is equal to $\sum_{i=1}^{n} (l_{2i} + l_{2i-1})(n - i + 1/2) + nl_0$.

From this version of the scheduling problem, we show how it can be reduced to our problem in the following. We set ϵ to be a very small constant value less than, say for example $\min\{l_0, \min_{i=0}^{2n-1}(l_{i+1} - l_i)/(100n^3)\}$. We set the map area to be $[0, 2\epsilon] \times [0, 2M]$ (see Figure 6).

We put all point sites in the map area along the vertical line $x = \epsilon$. For job $J_i, i = 0, 1, \cdots, 2n$, we introduce its corresponding point site p_i placing at location $(\epsilon, M - i\epsilon)$ such that the leader for p_i connects to the middle position of label L_i with height l_i (see Figure 6). Moreover, we set $k = \sum_{i=1}^{n}(l_{2i} + l_{2i-1})(n - i + 1/2) + nl_0$. Since ϵ is very small compared to any label height, we nearly can treat the locations of point sites p_i to be exactly lying at (ϵ, M). Under such a scenario, it can be shown that there is a scheduling with cost at most k if and only if there is a legal labeling with total leader length at most $k + l_0/n$. $\square$

Theorem 4. *It is NP-complete to find a* 1.5*-side boundary labeling of the model* (*non- uniform*, *FR*/*FP*/*sliding*, *TBM*)*.*

Proof (Sketch). We only discuss the sliding port case, because the fixed-ratio port case is similar. It is obvious to see that the problem is in NP. Hence, it suffices to show that the problem is NP-hard. The proof of the NP-completeness is based on the reduction from the following *subset sum* problem: Given $A = \{a_1, a_2, \cdots, a_n\}$ and a number B, the objective of the problem is to find a subset $A' \subset A$ such that the sum of the elements in A' is exactly B.

Given a set $A = \{a_1, \cdots, a_n\}$ and a number B, we construct a map instance as follows. There are $n+2$ sites p_1, p_2, $\ldots$, p_{n+2} in the map instance. The height of the label connected with site p_i is denoted by h_i, in which $h_{min} < \min\{h_1, \cdots, h_n\}$ and $h_{min} > n$. Let $\{h_1, \cdots, h_n\} = \{a_1, \cdots, a_n\}$, $h_{n+1} = h_{n+2} = k \leq h_{min}$, and $a_1 + \cdots + a_n = 2B$. Note that the y-coordinate of the bottom side of the map is zero. For each $i \in \{1, \cdots, n\}$, $y(p_i) \leq h_{min}$, $y(p_{n+1}) = h/2 - \epsilon$, and $y(p_{n+2}) = h/2 + \epsilon$, where h is the height of the map and $\epsilon < h_{min}$.

See also Figure 7. It is easy to see that the sum of the elements in A' is exactly $B = (a_1 + \cdots + a_n)/2$ if and only if the number of bends is $2n-2$. □

4.2 Pseudo-polynomial Time Algorithm

An idea which parallels the one used in Theorem 2 is used to show the following:

Theorem 5. *With respect to* 1.5*-side boundary labeling, both TLLM and TBM can be solved in* $O(n^4 h)$ *time when labels are of nonuniform height and ports are either fixed-ratio, fixed-position, or sliding.*

Proof. We only show how to solve the *TLLM* problem, since the *TBM* problem can be solved similarly. Like in the proof of Theorem 2, our pseudo-polynomial time algorithm is again based on the strategy of dynamic programming.

In the case of using fixed-ratio or fixed-position ports, we define $\gamma(p_i, t)$ (resp., $\ell(p_i, t)$), $1 \leq i \leq n, 0 \leq t < t' \leq h$, to be the length of a direct (resp., indirect) type-*opo* leader from site p_i to the fixed port of l_i when the y-coordinate of the bottom (resp., top) of the label is t (resp., t') in the label stack, provided that $t' - t = h_i$; if $t' - t \neq h_i$, then $\gamma(p_i, t)$ (resp., $\ell(p_i, t)) = \infty$. Note that t' is calculated in the dynamic programming procedure. In the case of using sliding ports, $\gamma(p_i, t)$ and $\ell(p_i, t)$ are defined similarly except that a leader connects p_i to the closest point (as opposed to the fixed point) on the boundary of l_i. We let $S(a, b, t)$ (where $1 \leq a \leq b \leq n, 0 \leq t \leq h$) denote the minimal total leader length where sites $p_a, \cdots, p_b$ ($1 \leq a \leq b \leq n$) are connected to their labels $l_a, \cdots, l_b$ which are placed in the label stack with t as the y-coordinate of the bottom of the lowest label. In other words, labels $l_a, \cdots, l_b$ occupy the area whose y-coordinate ranges from t to $t + \sum_{i=a}^{b} h_i$. We define $R_{a,b}$ to be the rectangle $[0, w] \times [y_{a-1}, y_{b+1}]$.

Our dynamic programming formula for $S(a, b, t)$ is as follows:

$$\min\{\sum_{i=0}^{b-a} \gamma(p_{a+i}, t + \sum_{j=0}^{i-1} h_{a+j}),$$

$$\min_{i,j\in\{0,\cdots,b-a\},j<i}\{\ell(p_{a+i},t+\sum_{l=0}^{j-1}h_{a+l})+S(a,a+j-1,t)$$
$$+S(a+j,a+i-1,t+(\sum_{l=0}^{j-1}h_{a+l})+h_{a+i})+S(a+i+1,b,t+\sum_{l=0}^{i}h_{a+l})\},$$
$$\min_{i,j\in\{0,\cdots,b-a\},j>i}\{\ell(p_{a+i},t+\sum_{l=0}^{j-1}h_{a+l})+S(a,a+i-1,t)$$
$$+S(a+i+1,a+j,t+\sum_{l=0}^{i-1}h_{a+l})+S(a+j+1,b,t+\sum_{l=0}^{j}h_{a+l})\}\}$$

The reason why the above dynamic programming formula correctly characterizes $S(a, b, t)$ is similar to that in the proof of Theorem 2. □

4.3 Fixed-Parameter Algorithm

We present a polynomial time algorithm when the number of different heights of labels, k, is a constant. First, we have the observation that the number of possible label positions is $O(n^k)$.

Lemma 2. *The number of possible label positions of each label is bounded by $O(n^k)$, if the given labels are of k different heights.*

Proof. We prove this lemma by induction on k. When $k = 1$, it is trivially true. Suppose that it is true for $k-1$, i.e., each label has $O(n^{k-1})$ possible label positions for n labels with $k-1$ different heights. Then it suffices to show that the lemma is true for k.

Since labels are placed without overlaps, we consider the location of the bottom side of a label as a label position. Consider each label L_i, which is the i-th label from the bottom. Let η_k denote the k-th kind of label height. Suppose that there are m labels of height η_k below label L_i, and hence, there are $(i-1-m)$ labels of the other $k-1$ kinds of heights below label L_i. Since the modification of the ordering of the labels below label L_k does change the position of L_i, we push all these m labels of height η_k down to the bottom and move other labels upwards accordingly. That is, we consider the label stack below label L_i: from the bottom there are m labels of height η_k and then $(i-1-m)$ labels of the other $k-1$ kinds of heights.

Note that $m < i \leq n$. For each possible $m = 1, 2, \cdots, i-1$, by inductive hypothesis, there are $O(n^{k-1})$ possible label positions for the $(i-1-m)$ labels of other kinds of heights. As a result, there are $O((i-1)\times n^{k-1}) = O(n^k)$ possible label positions for each label L_i. □

Since the $O(n^4 h)$ time algorithm in Theorem 5 considers h positions for the placement of labels, with Lemma 2, we see that the concerned problem can be solved in $O(n^4 \cdot n^k) = O(n^{k+4})$ time. The result is stated as follows.

Theorem 6. *With respect to* 1.5*-side boundary labeling, both TLLM and TBM can be solved in* $O(n^{k+4})$ *time when the labels are of constant* k *different heights and ports are either fixed-ratio, fixed-position, or sliding.*

5 Conclusion

We have investigated the total leader length minimization problem and the bend number minimization problem for one-and-a-half-side boundary labeling under a variety of settings parameterized by the underlying label size (uniform vs. nonuniform) and port type (fixed-ratio, fixed-position, vs. sliding). It turns out the two problems under the uniform-label case are solvable in polynomial time, whereas the problems become NP-complete under the nonuniform-label case. A pseudo-polynomial time algorithm and a fixed-parameter algorithm have been proposed for those intractable problems. In addition, the case where indirected leaders are not beneficial was also identified.

References

1. Bekos, M., Kaufmann, M., Nöllenburg, M., Symvonis, A.: Boundary labeling with octilinear leaders. Algorithmica 57(3), 436–461 (2010)
2. Bekos, M., Kaufmann, M., Potina, K., Symvonis, A.: Area-feature boundary labeling. The Computer Journal 53(6), 827–841 (2009)
3. Bekos, M., Kaufmann, M., Symvonis, A., Wolff, A.: Boundary labeling: models and efficient algorithms for rectangular maps. Computational Geometry: Theory and Applications 36(3), 215–236 (2006)
4. Benkert, M., Haverkort, H., Kroll, M., Nöllenburg, M.: Algorithms for multi-criteria one-sided boundary labeling. In: Hong, S.-H., Nishizeki, T., Quan, W. (eds.) GD 2007. LNCS, vol. 4875, pp. 243–254. Springer, Heidelberg (2008)
5. Formann, M., Wagner, F.: A packing problem with applications to lettering of maps. In: Proc. of the 7th Annual ACM Symposium on Computational Geometry (SoCG 1991), pp. 281–288. ACM Press, New York (1991)
6. Garey, M., Tarjan, R., Wilfong, G.: One-processor scheduling with symmetric earliness and tardiness penalties. Math. Oper. Res. 13, 330–348 (1988)
7. Lin, C.C.: Crossing-free many-to-one boundary labeling with hyperleaders. In: Proc. of 3rd IEEE Pacific Visualization Symposium (PacificVis 2010), pp. 185–192. IEEE Press, Los Alamitos (2010)
8. Lin, C.C., Kao, H.J., Yen, H.C.: Many-to-one boundary labeling. Journal of Graph Algorithms and Applications 12(3), 319–356 (2008)
9. Lin, C.C., Wu, H.Y., Yen, H.C.: Boundary labeling in text annotation. In: Proc. of 13th International Conference on Information Visualisation (IV 2009), pp. 110–115. IEEE CS Press, Los Alamitos (2009)
10. Wagner, F.: Approximate map labeling is in $\Omega(n \log n)$. Information Processing Letters 52(3), 161–165 (1994)
11. Wagner, F., Wolff, A.: Map labeling heuristics: Provably good and practically useful. In: Proc. of the 11th Annual ACM Symposium on Computational Geometry (SoCG 1995), pp. 109–118. ACM Press, New York (1995)