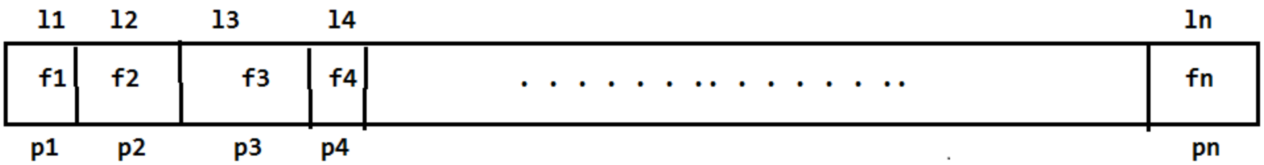
קורס אלגוריתמים 1. שעור 6 בעיית הקומפיילר.

ניסוח הבעיה. התוכנות שקומפיילר משתמש בהן שמורות על סרט מגנטי של זיכרון. לכל תוכנה יש שם $\boldsymbol{f_i}$ ונתון את אורכה $\boldsymbol{l_i}$ (ב-bytes) וההסתברות $\boldsymbol{p_i}$ שימוש בה. כדי להגיע לתוכנה כלשהי הקומפיילר צריך לעבור את כל התוכנות הנמצאות לפניה בסרט המגנטי.



המטרה: לסדר את התוכנות על הסרט המגנטי כך שזמן ריצה ממוצע יהיה נמוך ביותר. סה"כ יש $n!$ אפשרויות לסדר את הפונקציות על הסרט.

זמן ריצה ממוצע הוא:

$$T = l_1p_1 + (l_1 + l_2)p_2 + \cdots + (l_1 + \cdots + l_n)p_n$$

חיפוש שלם.

בחיפוש שלם צריך לעבור על כל הפרמוטציות האפשריות. לסדרה שמורכבת מ-n איברים

$(l_i, p_i), i = 1 \dots n$ מספר פרמוטציות הוא $n! > 2^n$ – לא יעיל לגרי.

נניח לרגע שכל התדירויות שוות: $p_1 = p_2 = \cdots = p_n = p$, במקרה זה

$$T = \big(l_1 + (l_1 + l_2) + \cdots + (l_1 + \cdots + l_n)\big)p$$

ומקבלים את בעיית המזכירה. במקרה זה כדי למזער את זמן ריצה הממוצע צריך למיין את המערך $\{l_i\}$ בסדר עולה (מקטן לגדול).

כאשר כל הפונקציות הן שווי-אורך, נקבל

$$\mathrm{T} = l(p_1 + 2p_2 + \cdots + np_n)$$

במקרה זה כדי למזער את זמן ריצה הממוצע צריך למיין את המערך $\{p_i\}$ בסדר יורד (מגדול לקטן).

במקרה הכללי ננסה להבין לפי מה נקבע את סדר התוכנות הנמצאות על הסרט המגנטי. לשם כך נחליף את שתי תכנות סמוכות: תוכנה הנמצאת במקום i עם תכנה הנמצאת במקום $i+1$:

$$T = l_1p_1 + (l_1 + \cdots + l_{i-1} + l_i)p_i + (l_1 + \cdots + l_i + l_{i+1})p_{i+1} + +(l_1 + \cdots + l_n)p_n$$

$$\bar{T} = l_1p_1 + (l_1 + \cdots + l_{i-1} + l_{i+1})p_{i+1} + (l_1 + \cdots + l_{i+1} + l_i)p_i + (l_1 + \cdots + l_n)p_n$$

$$T - \bar{T} = p_{i+1}l_i - p_il_{i+1}$$

אנו מקבלים זמן קטן יותר כאשר $T < \bar{T}$ או $T - \bar{T} > 0$, או

$$T - \bar{T} > 0 \;\rightarrow\; p_{i+1}l_i - p_il_{i+1} > 0 \;\rightarrow\; p_{i+1}l_i > p_il_{i+1} \;\rightarrow\; \frac{p_{i+1}}{l_{i+1}} > \frac{p_i}{l_i}$$

כלומר כדי לקבל זמן ממוצע קטן ביותר צריך למיין מערך של יחסים $\left\{\frac{p_i}{l_i}\right\}$ בסדר יורד מגדול לקטן. אמנם, אם יש זוג איברים סמוכים בסדרה $\left\{\frac{p_i}{l_i}\right\}$, כאשר הבא גדול מקודם ניתן להקטין את זמן הממוצע בעזרת שינוי הסדר של האיברים.

סיבוכיות האלגוריתם היא $O(nlog_2n) + O(n)$, כאשר $O(n)$ – סיבוכיות של חישוב הסדרה $\left\{\frac{p_i}{l_i}\right\}$ ו- $O(nlog_2n)$ סיבוכיות המיון.