הרצאה אלגוריתמים מתקדמים 28/12
מרצה: ד"ר יאיר ברטל

דפדוף (paging)

יש זיכרון בעל שתי רמות: זיכרון מהיר cache, M1, בעל גודל k וזיכרון איטי, M2 בעל גודל n>>k.

מגיעה סדרת גישות לדפים. כאשר יש בקשה לדף p:
אם p בזיכרון המהיר אז העלות היא אפס, אחרת (page fault) צריך להביא את p מהזיכרון M2 ל-M1 בעלות 1, ואם M1 היה מלא אז צריך לפנות דף אחר מ-M1 לפני שמביאים את p.

Offline (Sleator & Tarjan)

LFD – Longest Forward Distance
אלגוריתם אופטימלי: נפנה את הדף ב – M1 שיתבקש בעתיד הכי רחוק.
משפט: יש חסם תחתון של k על יחס התחרותיות של בעיית הדפדוף.
הוכחה: נבחר n=k+1. עבור סדרה σ של בקשות על k+1 דפים: $OPT(\sigma) \le \frac{|\sigma|}{k}$
נחלק את σ לפאזות שבכל אחת יש בקשות ל-k דפים שונים:
$$\sigma = \sigma^{(1)}, \sigma^{(2)}, ..., \sigma^{(k)}$$
כאשר $\sigma^{(i)}$: התת סדרה הקצרה ביותר ביותר שמכילה k דפים שונים.
OPT משרת את $\sigma^{(i)}$ ע"י הוצאת הדף הראשון של $\sigma^{(i+1)}$ והבאת הדף הראשון של $\sigma^{(i)}$. עלות $\sigma^{(i)}$ היא 1.
מצד שני, לכל אלגוריתם מכוון ON, ניתן להגדיר סדרה σ כך ש |σ|=ON(σ). היריב תמיד יבקש את הדף שהאלגוריתם ON הוציא.
ולכן ON(σ)≥k·OPT(σ).

<u>אלגוריתמים מעשיים טובים</u>

LRU – Least Recently Used, כשהזכרון מלא, הוצא את הדף שהתבקש בעבר הכי רחוק
FIFO – First In First Out, כשהזכרון מלא, הוצא את הדף שהוכנס ראשון
FWF – Flush When Full, כשהזכרון מלא, הוצא את כל הדפים

<u>אלגוריתם מעשיים רעים</u>

LIFO – Last In First Out, כשהזכרון מלא, הוצא את הדף שהוכנס אחרון

אלגוריתמי סימון, Marking Algorithm

מחלקה של אלגוריתמים k תחרותיים.
נגדיר חלוקה של הסידרה σ לפאזות. נגדיר את תהליך סימון הדפים ב – M1 הבא:

1. בהתחלת פאזה כל הדפים לא מסומנים
2. כאשר מבקשים דף במהלך פאזה הוא מסומן
3. בסיום פאזה מוחקים את כל הסימונים

האלגוריתם הוא marking אם בכל שלב בפאזה הוא מוציא רק דפים לא מסומנים.
__משפט:__ כל אלגוריתם marking הוא k תחרותי.
__הוכחה:__ במהלך פאזה יש בקשות ל – k דפים שונים. במקרה הגרוע הם לא היו ב – cache.
המחיר של אלגוריתם בפאזה $\leq k$. נראה שניתן לשייך לאלגוריתם היריב מחיר 1 לכל פאזה.
נתבונן בפאזה i, $\sigma^{(i)}$, ונניח ש – p הוא הדף הראשון שמבקשים בפאזה. במהלך $\sigma^{(i)}$
מבקשים k דפים שונים ונניח ש – q, הוא הדף הראשון שמבקשים ב - $\sigma^{(i+1)}$. אז בסה"כ יש
k+1 דפים שונים מהבקשה ל – p ועד הבקשה ל- q. כלומר, במהלך הזמן הזה, היתה עלות 1
לפחות ליריב וניתן לשייך את העלות הזו לפאזה ה – i.

__דוגמאות לאלגוריתמי סימון:__ FWF, LRU.
(FIFO אינו אלגוריתם סימון).

RANDOM MARKING

אלגוריתם סימון שמגריל בהסתברות אחידה איזה מבין הדפים הלא – מסומנים להוציא.
__משפט:__ RM הוא אלגוריתם סימון $2H_k$ תחרותי, כאשר

$$H_k = 1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{k}$$

כלומר, לכל קלט σ,

$$E\left[ON(\sigma)\right] \leq 2H_k \cdot OPT(\sigma)$$

__משפט:__ יש חסם תחתון של H_k על יחס התחרותיות של אלגוריתמים רנדומיים.
__הוכחה:__ נתבונן על חלוקה של σ לפאזות.

$$\sigma:\ \sigma^{(1)},\sigma^{(2)},\ldots$$

נגדיר:
דף "*ישן*" בפאזה i: דף שלא היה ב – cache בתחילת הפאזה.
דף "*חדש*" בפאזה i: דף שהתבקש בפאזה i ולא היה ב - $\sigma^{(i-1)}$ (לא ישן).
n_i את מספר הדפים החדשים בפאזה i.
ON_i: מחיר ON בפאזה i.
OPT_i: מחיר OPT בפאזה i.

__טענה (1):__ $E\left[ON(i)\right] \leq H_k n_i$

__טענה (2):__ $OPT(\sigma) \geq \frac{1}{2}\sum_{i=1}^{t} n_i$

מסקנה: $E\left(ON(\sigma)\right) \le H_k \sum_{i=1}^{t} n_i \le 2H_k OPT(\sigma)$

הוכחה (2): $OPT_{i-1}+OPT_i \ge n_i$. נובע מכך שב - $\sigma^{(i)}$ וב - $\sigma^{(i-1)}$ יש בקשות ל - $k+n_i$ דפים שונים

$$\Rightarrow \sum_{i=2}^{t}\left(OPT_{i-1}+OPT_i\right) \ge \sum_{i=2}^{t} n_i$$

$$-OPT_1 + 2\sum_{i=1}^{t} OPT_i \ge \sum_{i=2}^{t} n_i$$

$$2\sum_{i=1}^{t} OPT_i \ge OPT_1 + \sum_{i=2}^{t} n_i = \sum_{i=1}^{t} n_i$$

הוכחה (1): יש $k-n_i$ דפים ישנים שמתבקשים במהלך הפאזה ה – i. נסמן אותם: $p_1, p_2, \ldots, p_{k-n_i}$.

$$E\left(ON_i\right) = n_i + \sum_{j=1}^{k-n_i} \Pr\left(p_j \textit{ is not in the cache when it is asked for the first time}\right)$$

נסמן ב - $\tilde{n}_j$ את מספר הדפים החדשים שהתבקשו עד לבקשה הראשונה לדף p_j.

יש $j-1$ דפים ישנים שכבר התבקשו ואנחנו בוחנים את הבקשה ל – p_j. מה הסיכוי ש – p_j בין $\tilde{n}_j$ הדפים הראשונים שהוצאנו?

בסה"כ יש $k-(j-1)$ דפים ישנים.

מתוכם $\tilde{n}_j$ ב- M2 ו – $k-(j-1)-\tilde{n}_j$ ב – M1.

$\tilde{n}_j$ הדפים הישנים ב – M2 הם קבוצה אקראית מתוך הדפים הישנים שלא התבקשו (לא מסומנים).

מכאן שבזמן הבקשה,

$$\Pr\left(p_j \in M_2\right) = \frac{\tilde{n}_j}{k-(j-1)}$$

$$E\left(ON_i\right) = n_i + \sum_{j=1}^{k-n_i} \frac{\tilde{n}_i}{k-(j-1)} \underbrace{\le}_{\tilde{n}_j \le n_i} n_i + n_i \sum_{j=1}^{k-n_i} \frac{1}{k-(j-1)}$$

$$= n_i + n_i\left(\frac{1}{k}+\frac{1}{k-1}+\ldots+\frac{1}{n_i+1}\right) \underbrace{\le}_{n_i \ge 1} n_i + n_i\left(1+\frac{1}{2}+\ldots+\frac{1}{k}\right)$$