Analyse des Marker-Algorithmus

Hier ist der noch fehlende Schritt in der Analyse des Marker-Algorithmus. Erinnern wir uns: um die Kosten (Anzahl der cache misses) des Marker-Algorithmus in einer Runde abzuschätzen, haben wir die k Seiten, die zum ersten Mal in der Runde angefragt werden, in ℓ *neue* und $k-\ell$ *aufgefrischte* Seiten eingeteilt, je nachdem ob sie zu Beginn der Runde im cache (aufgefrischt) oder nicht im Cache (neu) waren.

Die Kosten für die ℓ Anfragen an neue Seiten sind demnach ℓ, weil jede solche Anfrage einen cache miss produziert. Die Kosten bei der Anfrage an die i-te aufgefrischte Seite der Runde hängen davon ab, ob diese Seite zum Anfragezeitpunkt im Cache (Kosten 0) oder nicht im Cache (Kosten 1) ist. Die *erwartete* Anzahl von cache misses bei der Anfrage ist also gleich der Wahrscheinlichkeit, dass die i-te aufgefrischte Seite zum Anfragzeitpunkt nicht im Cache ist. Bezeichne dieses Ereignis mit E_i.

Lemma 1.

$$\text{prob}(E_i) \leq \frac{\ell}{k-i-1}.$$

Beweis. Betrachte den Zeitpunkt der Anfrage an die i-te aufgefrischte Seite. Sei $c(i) \leq \ell$ die bisherige Anzahl von Anfragen an neue Seiten. Von den k Seiten, die zu Beginn der Runde im Cache waren, wurden bisher $i-1$ angefragt (nämlich genau die aufgefrischten), und diese sind im Cache. Von den $k-i+1$ nicht angefragten Seiten sind deshalb genau $c(i)$ nicht mehr im Cache, denn zur Zeit enthält der Cache ja $c(i)$ Seiten, die zu Beginn nicht im Cache waren. Ferner gilt, dass jede der $k-i+1$ noch nicht angefragten Seiten die gleiche Wahrscheinlichkeit p hat, nicht mehr im Cache zu sein, denn jedesmal, wenn eine dieser $k-i+1$ Seiten den Cache verlässt, so trifft es nach Konstruktion des Marker-Algorithmus jede der noch verbleibenden Seiten mit gleicher Wahrscheinlichkeit.

Folglich ist $p(k-i+1)$ die erwartete Anzahl der noch nicht angefragten initialen Cache-Seiten, die nicht mehr im cache sind. Andererseits ist diese (erwartete) Anzahl unabhängig vom Zufallsexperiment gleich $c(i)$, es gilt also

$$p = \frac{c(i)}{k-i+1}.$$

Betrachte nun die i-te aufgefrischte Seite. Diese ist eine der $k-i+1$ noch nicht angefragten Seiten, die zu Beginn im Cache waren und hat somit Wahrscheinlichkeit $p = \text{prob}(E_i)$, nicht mehr im Cache zu sein. Mit $c(i) \leq \ell$ folgt das Lemma.

$\square$