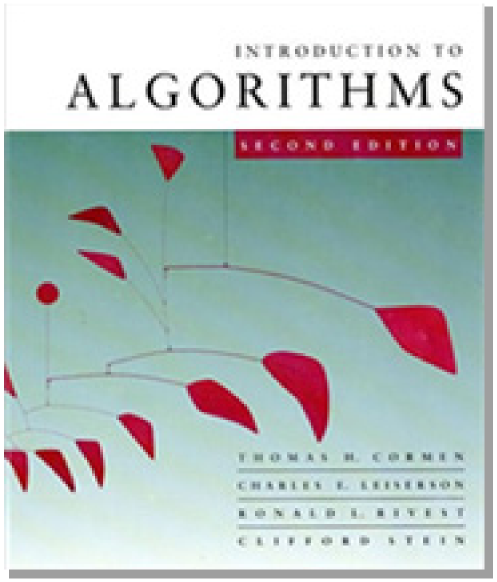
Algoritmalara Giriş

6.046J/18.401J

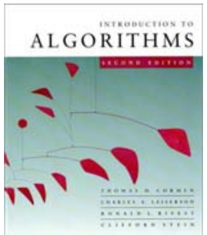


DERS 8

Kıyım Fonksiyonu(Hashing II)

- Evrensel kıyım fonksiyonu
- Evrensellik teoremi
- Evrensel kıyım fonksiyonları kümesini yapılandırmak
- Mükemmel kıyım fonksiyonu

Prof. Charles E. Leiserson



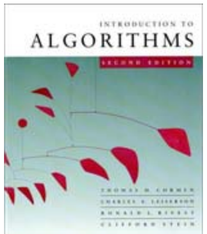
Kıyım fonksiyonunun bir zaafı

Problem: Her kıyım fonksiyonu h için, kıyım tablosuna ortalama erişim süresini çok büyük ölçüde arttıracak bir anahtar kümesi vardır.

- Rakibiniz bir i yuvası için tüm anahtarları $\{k \in U : h(k) = i\}$'den elde edebilir.

Fikir: Kıyım fonksiyonunu tüm anahtarlardan bağımsız olacak şekilde rastgele seçin.

- Rakibiniz kodunuzu görüyor olsa bile, hangi kıyım fonksiyonunun seçileceğini kesinlikle bilmediğinden, kötü bir anahtar kümesi bulamayacaktır.

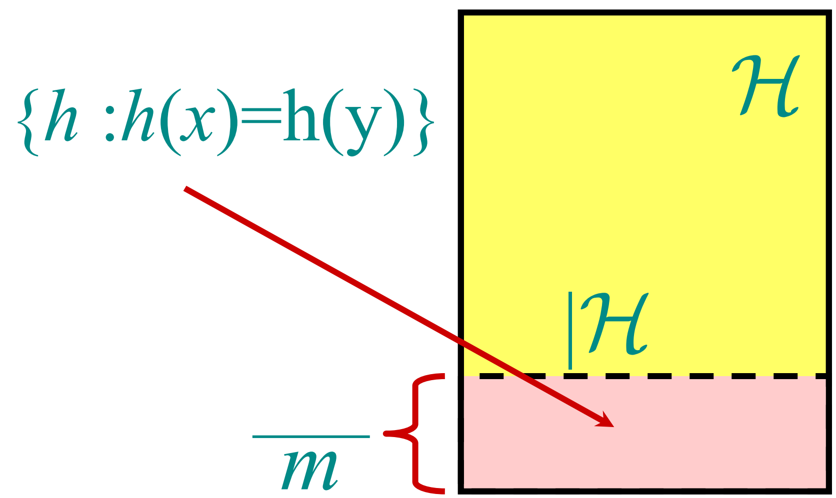


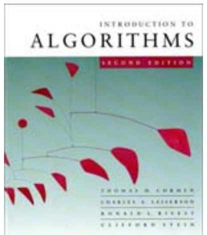
Evrensel kıyım fonksiyonu

Tanım. U bir anahtarlar evreni ve $\mathcal{H}$ de sınırlı sayıdaki kıyım fonksiyonlarının kümesi olsun; herbiri U ' yu $\{0, 1, \ldots, m-1\}$' e eşlemlesin.

$\mathcal{H}$ ' nin ***evrensel*** olması için: $x, y \in U$ ve $x \neq y$, ile $|\{h \in \mathcal{H} : h(x) = h(y)\}| = |\mathcal{H}|/m$ olması gerekir.

Yani, $\boldsymbol{x}$ ile $\boldsymbol{y}$ arasında bir çarpışma olasılığı : $1/m$ 'dir; koşul: h' nin $\mathcal{H}$' den rastgele seçimi.

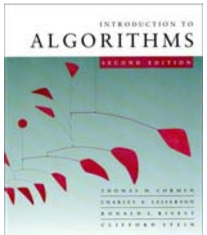




Evrensellik iyidir

Teorem. h (tekbiçimli olarak) rastgele seçilmiş bir kıyım fonksiyonu olsun; seçim evrensel bir $\mathcal{H}$ kıyım fonksiyonları setinden yapılmış olsun. h' nin n rastgele anahtarı T tablosundaki m yuvaya kıyımladığını farzedin.
Bu durumda verilen bir x anahtarı için:

$$E[x \text{ ile çarpışma sayısı}] < n/m.$$



Teoremin kanıtı

Kanıt. C_x, T ' nin içindeki anahtarlarla $\boldsymbol{x}$' in toplam çarpışma sayısını gösteren rastgele değişken olsun; ve

$$c_{xy} = \begin{cases} 1 \text{ eğer } h(x) = h(y), \\ 0 \text{ (diğer durumlarda)} \end{cases} \text{ olsun.}$$

Not: $E[c_{xy}] = 1/m$ ve $C_x = \displaystyle\sum_{y \in T - \{x\}} c_{xy}$.

Kanıt (devamı)

$$E[C_x] = E\left[\sum_{y \in T - \{x\}} c_{xy}\right]$$

- İki tarafın da beklenenini bulun.

Kanıt (devamı)

$$E[C_x] = E\left[\sum_{y \in T-\{x\}} c_{xy}\right]$$

$$= \sum_{y \in T-\{x\}} E[c_{xy}]$$

- İki tarafın da beklenenini bulun.
- Beklenenin doğrusallığı (expectation).

Kanıt (devamı)

$$E[C_x] = E\left[\sum_{y \in T - \{x\}} c_{xy}\right]$$

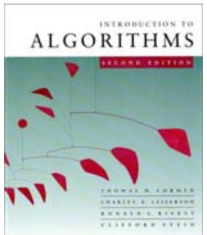
- İki tarafın da beklenenini bulun.

$$= \sum_{y \in T - \{x\}} E[c_{xy}]$$

- Beklenenin doğrusallığı (expectation).

$$= \sum_{y \in T - \{x\}} 1/m$$

- $E[c_{xy}] = 1/m$.



Kanıt (devamı)

$$
\begin{aligned}
E[C_x] &= E\left[\sum_{y \in T-\{x\}} c_{xy}\right] \\
&= \sum_{y \in T-\{x\}} E[c_{xy}] \\
&= \sum_{y \in T-\{x\}} 1/m \\
&= \frac{n-1}{m} . \quad \square
\end{aligned}
$$

- Her iki tarafın da beklenenini bulun.
- Beklenenin doğrusallığı (expectation).
- $E[c_{xy}] = 1/m$.
- Cebir.

Bir evrensel kıyım fonksiyonları setini yapılandırmak

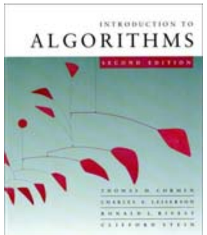
m asal sayı olsun. k anahtarını $r + 1$ basamağa ayrıştırın; herbirinin set içinde değeri $\{0, 1, \ldots, m-1\}$ olsun. Yani, $k = \langle k_0, k_1, \ldots, k_r \rangle$ ve $0 \le k_i < m$ olsun.

Rastgele yapma stratejisi:

$a = \langle a_0, a_1, \ldots, a_r \rangle$ olsun; burada a_i , $\{0, 1, \ldots, m-1\}$ arasından rastgele seçilmiştir.

Tanım: $h_a(k) = \sum_{i=0}^{r} a_i k_i \bmod m$. ***Nokta çarpım, mod m*** (ölçke)

$\mathcal{H} = \{h_a\}$ ne büyüklükte? $|\mathcal{H}| = m^{r+1}$. ← **BUNU HATIRLAYIN!**



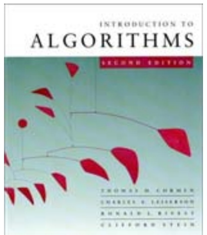
Nokta - çarpım kıyım fonksiyonların evrenselliği

Teorem. $\mathcal{H} = \{h_a\}$ seti evrenseldir.

Kanıt. $x = \langle x_0, x_1, \ldots, x_r \rangle$ olduğunu varsayın ve $y = \langle y_0, y_1, \ldots, y_r \rangle$ farklı anahtarlar olsun. Yani, en az bir basamakta farklı olsunlar ve log pozisyonu 0 olsun. Kaç $h_a \in \mathcal{H}$ için x ve y çarpışırlar?

$h_a(x) = h_a(y)$ olması gerekir ve bunun anlamı:

$$\sum_{i=0}^{r} a_i x_i \equiv \sum_{i=0}^{r} a_i y_i \pmod{m}$$



Kanıt (devamı)

Benzer yaklaşımla, elimizde

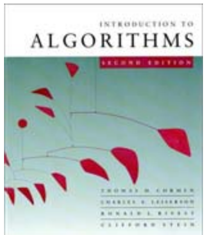
$$\sum_{i=0}^{r} a_i (x_i - y_i) \equiv 0 \pmod{m}$$

veya

$$a_0 (x_0 - y_0) + \sum_{i=1}^{r} a_i (x_i - y_i) \equiv 0 \pmod{m} \; ,$$

olur ve bu da şu anlama gelir:

$$a_0 (x_0 - y_0) \equiv -\sum_{i=1}^{r} a_i (x_i - y_i) \pmod{m} \; .$$



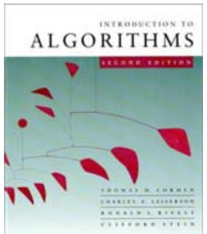
Sayı teorisinin gerçeği

Teorem. m asal sayı olsun. Herhangi bir $z \in \mathbb{Z}_m$ ve $z \neq 0$ için, özgün bir $z^{-1} \in \mathbb{Z}_m$ vardır ve bu durumda:

$$z \cdot z^{-1} \equiv 1 \pmod{m}.$$ (ölçke m) olur.

Örnek: $m = 7$.

z	1	2	3	4	5	6
z^{-1}	1	4	5	2	3	6



Kanıta geri dönüş

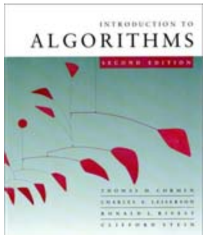
Elimizde

$$a_0(x_0 - y_0) \equiv -\sum_{i=1}^{r} a_i(x_i - y_i) \pmod{m} \text{ var,}$$

ve $x_0 \neq y_0$, olduğundan tersi de $(x_0 - y_0)^{-1}$ olmalıdır, ve bu da şu anlama gelir:

$$a_0 \equiv \left(-\sum_{i=1}^{r} a_i(x_i - y_i)\right) \cdot (x_0 - y_0)^{-1} \pmod{m}.$$

Yani, herhangi bir $a_1, a_2, \ldots, a_r$, seçiminde tek a_0 seçimi, x ile y 'nin çarpışmasına neden olur.



Kanıt (tamamlanması)

S. Kaç tane h_a, x ile y' nin çarpışmasına neden olur?

C. Her $a_1, a_2, \ldots, a_r$ için m seçenek vardır ama, bunlar bir kez seçildiğinde, sadece bir tane a_0 x ile y' yi çarpıştırabilir, yani

$$a_0 = \left(\left(-\sum_{i=1}^{r} a_i(x_i - y_i)\right) \cdot (x_0 - y_0)^{-1}\right) \bmod m .$$

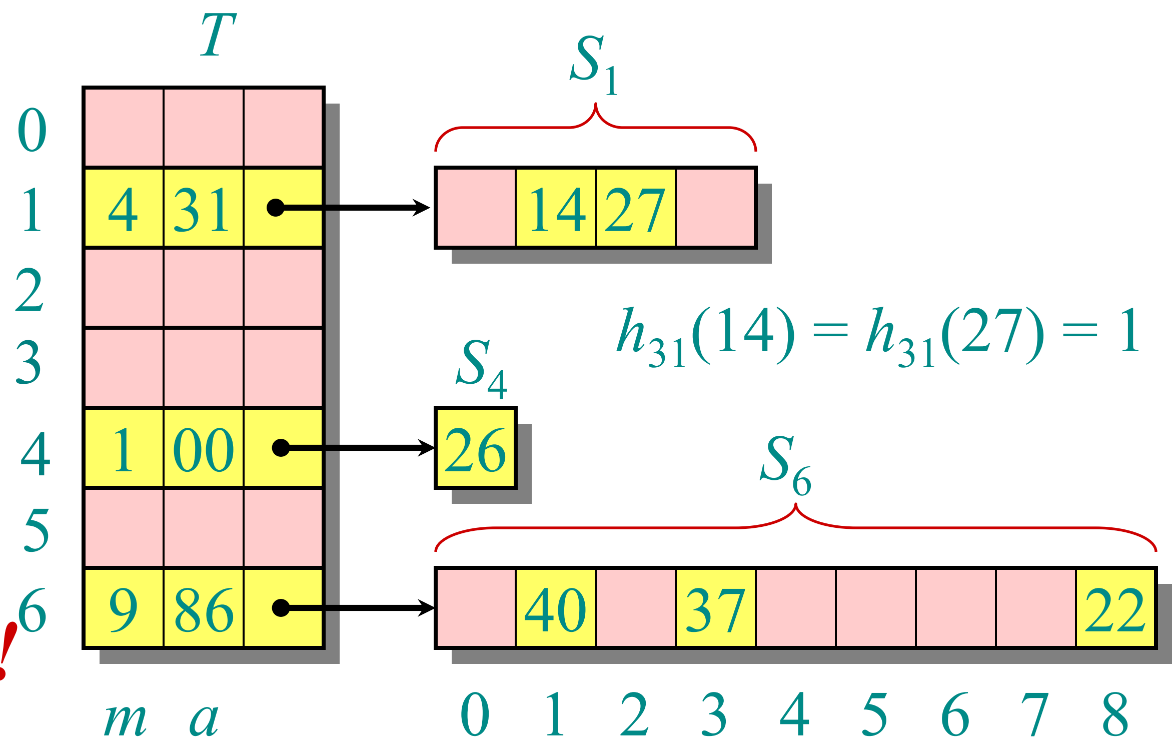
Böylece, çarpışmaya neden olabilecek h' lerin sayısı $= m^r \cdot 1 = m^r = |\mathcal{H}|/m$. □

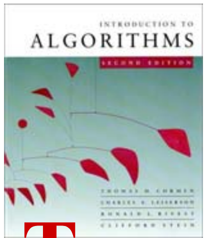
Mükemmel kıyım fonksiyonu

n anahtarlı bir set verilirse, bir statik kıyım tablosunu boyutu $m = O(n)$ olacak şekilde yapılandırın ve ARAMA (SEARCH) *en kötü durumda* $\Theta(1)$ süre alsın.

FİKİR: Her iki düzeyde de evrensel kıyım ile 2 düzeyli veri tanımlama.

2. düzey çarpışması yok!



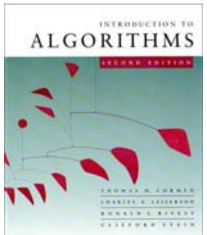


2. düzeyde çarpışmalar.

Teorem. $\mathcal{H}$, evrensel kıyım fonksiyonu türlerinden biri olsun ve boyutu da $m = n^2$ olsun. Bu durumda, eğer bir rastgele $h \in \mathcal{H}$ ' yi n anahtarı tabloya kıyımlamakta kullanırsak, beklenen çarpışma sayısı en çok $1/2$ olur.

Kanıt. Evrenselliğin tanımı gereği, tablodaki belirli 2 anahtarın h altında çarpışma olasılığı $1/m = 1/n^2$ olur. Çarpışma olasılıklı $\binom{n}{2}$ çift anahtar olduğundan, çarpışmaların beklenen sayısı:

$$\binom{n}{2} \cdot \frac{1}{n^2} = \frac{n(n-1)}{2} \cdot \frac{1}{n^2} < \frac{1}{2}. \quad \square$$



2. düzeyde çarpışma yoktur

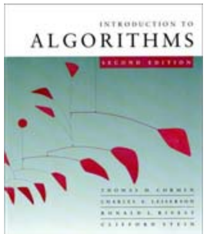
Corollary/ Doğal sonuç. Hiç çarpışma olmaması olasılığı en az $1/2$ dir.

Kanıt. ***Markov'un eşitsizliği*** çerçevesinde herhangi bir negatif olmayan rastgele değişken X için,

$$\Pr\{X \geq t\} \leq E[X]/t \text{ dir.}$$

Bu eşitsizliği $t = 1$, durumuna uygularsak, 1 yada daha fazla çarpışma olasılığının en çok $1/2$ olduğunu buluruz. $\square$

Böylece, $\mathcal{H}$ *'nin içindeki rastgele kıyım fonksiyonlarını test ederek, çabucak çalışan bir tanesini buluruz .*



Depolamanın çözümlenmesi

Düzey-1 değerli kıyım tablosu T için, $m = n$ seçin ve n_i de T'deki i yuvasına kıyımlanan anahtarları belirten rastgele değişken olsun. Burada n_i^2 yuvalı düzey-2 kıyım tablosu S_i kullanılırsa iki-düzeyli veri tanımlama işlemi için gerekli beklenen toplam depolama

$$E\left[\sum_{i=0}^{m-1} \Theta\left(n_i^2\right)\right] = \Theta(n) \text{ olur,}$$

çünkü buradaki çözümleme daha önce ele alınan sepet sıralamasının beklenen koşma süresindekinin aynıdır. (Olasılık sınırı için Markov'u uygulayın.)