# Dynamické pole s přidáváním prvků na konec

Máme datovou strukturu dynamického pole s jedinou operací Insert, která přidává prvky na konec. Dynamické pole si drží dva údaje – velikost (počet prvků, *size*) a kapacitu (rozměr alokované paměti, *capacity*). Na počátku nemá pole alokovánu žádnou paměť – jeho velikost i kapacita je 0. Operace Insert funguje následovně:

- Je-li kapacita rovna 0, alokuj paměť o kapacitě 1.
- Je-li aktuální velikost rovná kapacitě, alokuj novou paměť o dvojnásobné kapacitě, přesuň do ní všechny prvky ze staré paměti a dealokuj starou paměť.
- Vlož prvek na konec pole a zvětši velikost pole o 1.

Pro velikost a kapacitu dynamického pole $P$ budeme dále používat notaci $\mathit{size}(P)$ a $\mathit{capacity}(P)$.

Cena operace Insert je dána následovně (cena je dána pouze přesuny prvků, alokaci/dealokaci paměti do složitosti pro zjednodušení nepočítáme):

- Pokud $\mathit{size}(P) = \mathit{capacity}(P)$, pak Insert stojí $\mathit{size}(P) + 1$.
- Jinak Insert stojí 1.

Dokážeme nyní metodou potenciálové funkce, že amortizovaná složitost Insert je konstantní, tedy že libovolná posloupnost operací Insert délky $n$ má složitost v $\mathcal{O}(n)$.

Vhodná potenciálová funkce musí být dostatečně vysoká před provedením „drahého“ Insertu a dostatečně nízká po něm. Zároveň nesmí nikdy klesnout pod svou počáteční hodnotu.

Zvolíme potenciálovou funkci $\Phi(P) = 2 \cdot \mathit{size}(P) - \mathit{capacity}(P)$. Počáteční hodnota této potenciálové funkce je 0; zároveň je jasné, že nikdy v průběhu vykonávání Insertů nebude hodnota potenciálové funkce menší než 0, protože je implementací operace Insert zaručeno, že vždy platí $\mathit{size}(P) \geq \mathit{capacity}(P)/2$.

Zbývá tedy spočítat amortizovanou cenu $i$. operace Insert, pro všechna $i$ od 1 do $n$. Budeme používat následující notaci: $P_i$ značí stav dynamického pole *po* provedení $i$. operace. Počáteční stav pole je tedy označen $P_0$.

**„Levný“ Insert**: Platí $\mathit{size}(P_{i-1}) < \mathit{capacity}(P_{i-1})$ a skutečná cena Insertu je $c_i = 1$. Po provedení Insertu pak víme, že $\mathit{capacity}(P_i) = \mathit{capacity}(P_{i-1})$ a $\mathit{size}(P_i) = \mathit{size}(P_{i-1}) + 1$. Pak platí:

$$\begin{aligned}
\hat{c}_i &= c_i + \Phi(P_i) - \Phi(P_{i-1}) \\
&= 1 + (2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})) \\
&= 1 + (2 \cdot (\mathit{size}(P_{i-1}) + 1) - \mathit{capacity}(P_{i-1})) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})) \\
&= 1 + 2 = 3
\end{aligned}$$

**„Drahý“ Insert**: Platí $\mathit{size}(P_{i-1}) = \mathit{capacity}(P_{i-1})$ a skutečná cena Insertu je $c_i = \mathit{size}(P_{i-1}) + 1$. Po provedení Insertu pak víme, že $\mathit{capacity}(P_i) \geq 2 \cdot \mathit{capacity}(P_{i-1})$[1] a $\mathit{size}(P_i) = \mathit{size}(P_{i-1}) + 1$. Pak platí:

$$\begin{aligned}
\hat{c}_i &= c_i + \Phi(P_i) - \Phi(P_{i-1}) \\
&= (\mathit{size}(P_{i-1}) + 1) + (2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})) \\
&\leq (\mathit{size}(P_{i-1}) + 1) + (2 \cdot (\mathit{size}(P_{i-1}) + 1) - 2 \cdot \mathit{capacity}(P_{i-1})) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})) \\
&= \mathit{size}(P_{i-1}) + 1 + 2 - \mathit{capacity}(P_{i-1}) = 3
\end{aligned}$$

Ve všech případech je tedy $\hat{c}_i$ shora omezeno konstantou 3, a proto $\sum_{i=1}^{n} c_i \leq \sum_{i=1}^{n} \hat{c}_i \leq 3 \cdot n \in \mathcal{O}(n)$.

# Dynamické pole s odebíráním prvků z konce

## Nejjednodušší varianta bez zmenšování

Nejjednodušší možnost je při mazání pole nijak nezmenšovat. Operace Delete je pak vždy zaručeně konstantní, přičemž složitost operace Insert se nijak nezhorší. Striktně vzato bychom ale při analýze

[1] To zahrnuje jak případ s $\mathit{capacity}(P_{i-1}) = 0$, kdy $\mathit{capacity}(P_i) = 1$, tak případ s $\mathit{capacity}(P_{i-1}) > 0$, kdy $\mathit{capacity}(P_i) = 2 \cdot \mathit{capacity}(P_{i-1})$. Všimněte si, že z tohoto důvodu se v následujícím výpočtu objeví $\leq$.

amortizované složitost nemohli použít dříve uvedenou potenciálovou funkci (protože už nebude nutně platit $\textit{size}(P) \geq \textit{capacity}(P)/2$). Vhodná potenciálová funkce je v tomto případě tato:

$$\Phi(P) = \begin{cases} 2 \cdot \textit{size}(P) - \textit{capacity}(P) & \text{pokud je } \textit{size}(P) \geq \textit{capacity}(P)/2 \\ 0 & \text{jinak} \end{cases}$$

Výpočet $\hat{c}_i$ ve všech případech (je třeba uvažovat obě operace!) necháváme čtenáři jako cvičení.

## Špatná varianta se zmenšováním

Uvažujme nyní následující možnost: Delete smaže poslední prvek a pokud tím zaplnění dynamického pole kleslo pod polovinu, alokuje se nová paměť polovičního rozměru a prvky se do ní přesunou. Ukážeme, že tato varianta nemá dobrou amortizovanou složitost; konkrétně ukážeme, že existuje posloupnost $n$ operací Insert a Delete taková, jejíž celková složitost je v $\Omega(n^2)$. Skutečná cena Delete je při zmenšování pole rovna $\textit{size}(P)$.

Nechť $k$ je nejbližší mocnina dvou menší než nebo rovna $n/2$, tj. $k = 2^i$ pro nějaké přirozené $i$ takové, že $2^i \leq n/2 < 2^{i+1} = 2 \cdot k$. Posloupnost operací sestavíme následovně:

- Nejprve $k$krát provedeme operaci Insert.
- Následně $\lfloor (n-k)/4 \rfloor$krát provedeme čtveřici Insert, Delete, Delete, Insert.
- Pokud ještě nemáme posloupnost délky $n$, doplníme ji do délky $n$ operacemi Insert (tyto operace budou maximálně 3).

Je zřejmé, že po $k$ provedeních operace Insert bude dynamické pole mít velikost i kapacitu rovnu $k$. Cenu těchto operací ignorujeme (snažíme se nyní určit *dolní odhad* složitosti, takže si to můžeme dovolit).

Cena každé čtveřice v druhé části posloupnosti je následující: $k+1$ za první Insert, 1 za první Delete, $k$ za druhý Delete (přesun všech prvků), 1 za druhý Insert. Dohromady alespoň $\Omega(k)$. Takovýchto čtveřic je v naší posloupnosti $\lfloor (n-k)/4 \rfloor$. Dolní odhad pro tento počet je:

$$\lfloor (n-k)/4 \rfloor \geq (n-k)/4 \geq (n-(n/2))/4 = n/8.$$

Celková cena posloupnosti je tedy alespoň $k \cdot (n/8)$. Protože $n/2 < 2 \cdot k$, platí $n/4 < k$ a cena posloupnosti je tedy alespoň $(n/4) \cdot (n/8)$, což patří do $\Omega(n^2)$.

**Poznámka:** Všimněte si, že i v tomto případě je operace Insert amortizovaně konstantní, pokud bychom ji uvažovali odděleně. Stejně tak operace Delete je amortizovaně konstantní, pokud bychom ji uvažovali odděleně (např. tak, že začínáme s dostatečně velkým polem, na němž voláme pouze posloupnost operací Delete). Je zde tedy vidět, že máme-li víc operací, není možné určovat jejich amortizovanou složitost „po částech“, ale je třeba je uvažovat všechny zároveň.

## Správná varianta se zmenšováním

Uvažujme nyní následující variantu: Delete smaže poslední prvek a pokud tím zaplnění dynamického pole klesne přesně na *čtvrtinu*, alokujeme novou paměť *polovičního* rozměru a přesuneme prvky do nové paměti. Pole o kapacitě menší než 4 (tj. 1 nebo 2) nebudeme zmenšovat vůbec. Cena operace Delete je $\textit{size}(P)$, pokud $\textit{capacity}(P) \geq 4$ a $\textit{size}(P) = \textit{capacity}(P)/4 + 1$, jinak je 1.

Dokážeme nyní metodou potenciálové funkce, že amortizovaná složitost obou operací Insert i Delete je konstantní, tedy že libovolná posloupnost operací Insert a Delete délky $n$ má složitost v $\mathcal{O}(n)$.

Vhodná potenciálová funkce musí být dostatečně vysoká před provedením jakékoli „drahé“ operace (tj. když je pole plné nebo když se jeho zaplnění blíží čtvrtině) a dostatečně nízká po provedení této operace (tj. když je pole zaplněné do poloviny).

Potenciálová funkce uvedená v první části tohoto dokumentu splňuje jen část těchto požadavků. Zkusíme to vyřešit tím, že ji zavřeme do absolutní hodnoty. Definujme tedy $\Phi(P) = |2 \cdot \textit{size}(P) - \textit{capacity}(P)|$. Počáteční hodnota této funkce je 0; že nikdy neklesne pod 0 je dáno vlastností absolutní hodnoty.

Zbývá spočítat amortizovanou cenu $i$. operace (což může být Insert nebo Delete), pro všechna $i$ od 1 do $n$. Používáme stejnou notaci jako v prvním příkladu. Je třeba uvážit všechny možnosti.

Nezapomínejte, že ukazujeme *horní odhad* složitosti a stačí nám tedy, když bude amortizovaná cena $\hat{c}_i$ *shora ohraničená námi požadovanou funkcí* (v tomto případě konstantou).

**„Levný" Insert**: Platí $\mathit{size}(P_{i-1}) < \mathit{capacity}(P_{i-1})$ a skutečná cena Insertu je $c_i = 1$. Po provedení Insertu víme, že $\mathit{capacity}(P_i) = \mathit{capacity}(P_{i-1})$ a $\mathit{size}(P_i) = \mathit{size}(P_{i-1}) + 1$. Pak platí:

$$\begin{aligned}\hat{c}_i &= c_i + \Phi(P_i) - \Phi(P_{i-1})\\ &= 1 + |2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= 1 + |2 \cdot (\mathit{size}(P_{i-1}) + 1) - \mathit{capacity}(P_{i-1})| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= 1 + |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}) + 2| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\end{aligned}$$

V tuto chvíli je třeba další výpočet rozvětvit podle toho, jak se realizují absolutní hodnoty v našem výrazu. První možnost: $\mathit{size}(P_{i-1}) \geq \mathit{capacity}(P_{i-1})/2$. V tom případě pak jsou výrazy uvnitř obou absolutních hodnot nezáporné a můžeme pokračovat takto:

$$\begin{aligned}\hat{c}_i &= 1 + (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}) + 2) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}))\\ &= 1 + 2 = 3\end{aligned}$$

Druhá možnost: $\mathit{size}(P_{i-1}) < \mathit{capacity}(P_{i-1})/2$. V tom případe jsou výrazy uvnitř obou absolutních hodnot nekladné (zde je důležité si uvědomit, že $\mathit{size}(P_{i-1})$ je celé číslo) a můžeme pokračovat takto (všimněte si výměny znamének):

$$\begin{aligned}\hat{c}_i &= 1 - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}) + 2) + (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}))\\ &= 1 - 2 = -1\end{aligned}$$

Vadí nám nějak, že nám vyšlo záporné číslo? Nevadí, stále je to konstanta.

**„Drahý" Insert**: Platí $\mathit{size}(P_{i-1}) = \mathit{capacity}(P_{i-1})$ a skutečná cena Insertu je $c_i = \mathit{size}(P_{i-1}) + 1$. Po provedení Insertu pak víme, že $\mathit{capacity}(P_i) \geq 2 \cdot \mathit{capacity}(P_{i-1})$ (viz poznámku v prvním příkladu) a $\mathit{size}(P_i) = \mathit{size}(P_{i-1}) + 1$. Pak platí:

$$\begin{aligned}\hat{c}_i &= c_i + \Phi(P_i) - \Phi(P_{i-1})\\ &= (\mathit{size}(P_{i-1}) + 1) + |2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= (\mathit{size}(P_{i-1}) + 1) + (2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}))\\ &\leq (\mathit{size}(P_{i-1}) + 1) + (2 \cdot (\mathit{size}(P_{i-1}) + 1) - 2 \cdot \mathit{capacity}(P_{i-1})) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}))\\ &= \mathit{size}(P_{i-1}) + 1 + 2 - \mathit{capacity}(P_{i-1}) = 3\end{aligned}$$

Krok od druhého k třetímu řádku si můžeme dovolit, protože víme, že $\mathit{size}(P_{i-1}) \geq \mathit{capacity}(P_{i-1})/2$ a totéž platí pro $P_i$.

**„Levný" Delete:** Platí buď $\mathit{capacity}(P_{i-1}) < 4$ nebo $\mathit{size}(P_{i-1}) > \mathit{capacity}(P_{i-1})/4$. Skutečná cena Delete je 1. Po provedení Delete pak víme, že $\mathit{capacity}(P_i) = \mathit{capacity}(P_{i-1})$ a $\mathit{size}(P_i) = \max(\mathit{size}(P_{i-1}) - 1, 0)$ (z prázdného pole zřejmě není možno nic odebrat). Nejdříve vyřešíme situaci $\mathit{capacity}(P_{i-1}) < 4$:

$$\begin{aligned}\hat{c}_i &= c_i + \Phi(P_i) - \Phi(P_{i-1})\\ &= 1 + |2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\end{aligned}$$

Zřejmě první absolutní hodnota nemůže být větší než 3 (dá se snadno ověřit vyzkoušením všech možných hodnot $\mathit{size}(P_i) \leq \mathit{capacity}(P_i) < 4$). Zároveň druhá absolutní hodnota nemůže být menší než 0 (to je vlastnost absolutní hodnoty). Proto můžeme náš výraz omezit takto:

$$\hat{c}_i \leq 1 + 3 - 0 = 4$$

(Jistě, přesnější výpočet by nám tento výraz omezil více, ale proč? Počítáme *asymptotickou* složitost, tedy nám nezáleží na velikosti konstanty.)

Dále vyřešíme situaci $\mathit{capacity}(P_{i-1}) \geq 4$. Nyní jistě víme, že $\mathit{size}(P_{i-1}) = \mathit{size}(P_{i-1}) - 1$.

$$\begin{aligned}\hat{c}_i &= c_i + \Phi(P_i) - \Phi(P_{i-1})\\ &= 1 + |2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= 1 + |2 \cdot (\mathit{size}(P_{i-1}) - 1) - \mathit{capacity}(P_{i-1})| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= 1 + |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}) - 2| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\end{aligned}$$

Podobně jako výše u „levného“ Insertu, musíme nyní výpočet rozvětvit podle toho, jak se realizují absolutní hodnoty. První možnost: $\mathit{size}(P_{i-1}) > \mathit{capacity}(P_{i-1})/2$. V tom případě jsou výrazy uvnitř obou absolutních hodnot nezáporné:

$$\begin{aligned}\hat{c}_i &= 1 + (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}) - 2) - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}))\\ &= 1 - 2 = -1\end{aligned}$$

Druhá možnost: $\mathit{size}(P_{i-1}) \leq \mathit{capacity}(P_{i-1})/2$. V tom případě jsou výrazy uvnitř obou absolutních hodnot nekladné:

$$\begin{aligned}\hat{c}_i &= 1 - (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}) - 2) + (2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1}))\\ &= 1 + 2 = 3\end{aligned}$$

**„Drahý“ Delete**: Platí $\mathit{capacity}(P_{i-1}) \geq 4$, $\mathit{size}(P_{i-1}) = \mathit{capacity}(P_{i-1})/4 + 1$ a skutečná cena Delete je $\mathit{size}(P_{i-1})$. Víme, že po provedení Delete platí $\mathit{size}(P_i) = \mathit{size}(P_{i-1}) - 1 = \mathit{capacity}(P_{i-1})/4$ a $\mathit{capacity}(P_i) = \mathit{capacity}(P_{i-1})/2$.

$$\begin{aligned}\hat{c}_i &= c_i + \Phi(P_i) - \Phi(P_{i-1})\\ &= \mathit{size}(P_{i-1}) + |2 \cdot \mathit{size}(P_i) - \mathit{capacity}(P_i)| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= \mathit{size}(P_{i-1}) + |2 \cdot \mathit{capacity}(P_{i-1})/4 - \mathit{capacity}(P_{i-1})/2| - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= \mathit{size}(P_{i-1}) + 0 - |2 \cdot \mathit{size}(P_{i-1}) - \mathit{capacity}(P_{i-1})|\\ &= (\mathit{capacity}(P_{i-1})/4 + 1) - |2 \cdot (\mathit{capacity}(P_{i-1})/4 + 1) - \mathit{capacity}(P_{i-1})|\\ &= \mathit{capacity}(P_{i-1})/4 + 1 - |2 - \mathit{capacity}(P_{i-1})/2|\end{aligned}$$

Výraz uvnitř zbývající absolutní hodnoty je nekladný (víme, že $\mathit{capacity}(P_{i-1}) \geq 4$) a proto pokračujeme takto:

$$\begin{aligned}\hat{c}_i &= \mathit{capacity}(P_{i-1})/4 + 1 + (2 - \mathit{capacity}(P_{i-1})/2)\\ &= 3 - \mathit{capacity}(P_{i-1})/4 \leq 3\end{aligned}$$

Všimněte si, že nám nijak nevadí, že nám vyšel výraz, který je závislý na kapacitě dynamického pole. Podstatné je, že tento výraz umíme *shora omezit* konstantou, v tomto případě 3.

Ve všech případech je tedy $\hat{c}_i$ shora omezeno konstantou 4, a proto $\sum_{i=1}^{n} c_i \leq \sum_{i=1}^{n} \hat{c}_i \leq 4 \cdot n \in \mathcal{O}(n)$. (Ano, pokud bychom byli přesnější, mohlo nám vyjít $3 \cdot n$, ale zajímá nás *asymptotická* složitost.)