Detekcia vzorov správania procesu v postupnosti adries

Školiteľ: Ing. Dušan Bernát, PhD.

Matúš Hluch

Pamäťové vzory

- Pristupované adresy – vzory (Denning, 1972)

- Cykly, polia, dátové štruktúry, inštrukcie uložené v pamäti za sebou

- Sekvenčná lokalita – proces pristupuje na za sebou idúce adresy
- Časová lokalita – proces bude pristupovať na rovnakú adresu v blízkom čase

Motivácia

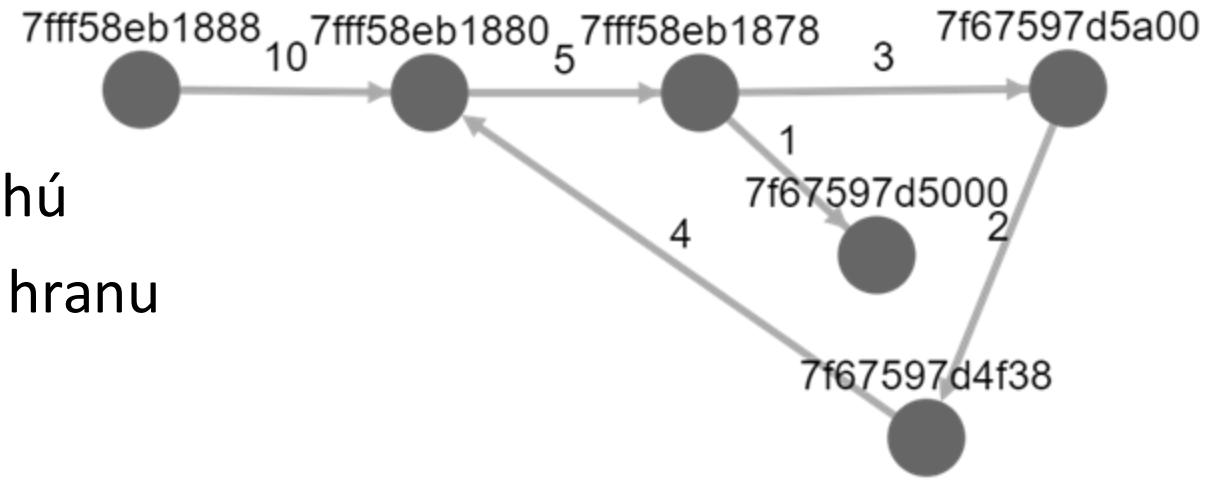
- Každý proces – obrovské počty pristúpených adries
  - Príkaz *ls* okolo 300 000
- Pamäťová hierarchia
- Denningove práce, 1972
- Najnovší výskum – bezpečnosť, vírusy

Analýza prístupov do pamäte

- Priame hľadanie vzorov v postupnosti adries

- Reprezentácia postupnosti adries ako ohodnotený orientovaný graf

- Vrcholy – adresy
- Hrany – prechody z jednej adresy na druhú
- Váha hrany – počet prechodov cez danú hranu



- Analýza štruktúry grafov

Vstupné dáta

- X86-64 architektúra, ARM 32-bit architektúra
- Oddelená inštrukčná a dátová pamäť
  - Cache pamäť
- Linuxové príkazy: *ls*, *cd*, *mkdir*, *wget*...
- Python, C

Pin a Dynamorio

- Existujúce nástroje určené na dynamickú analýzu programov
- Just-in-time kompilácia, pridávanie kódu do vykonávaného programu
- Pin X86-64, Dynamorio: ARM, x86...
- Záznam pristupovaných adries
  - Inštrukčná/Dátová pamäť

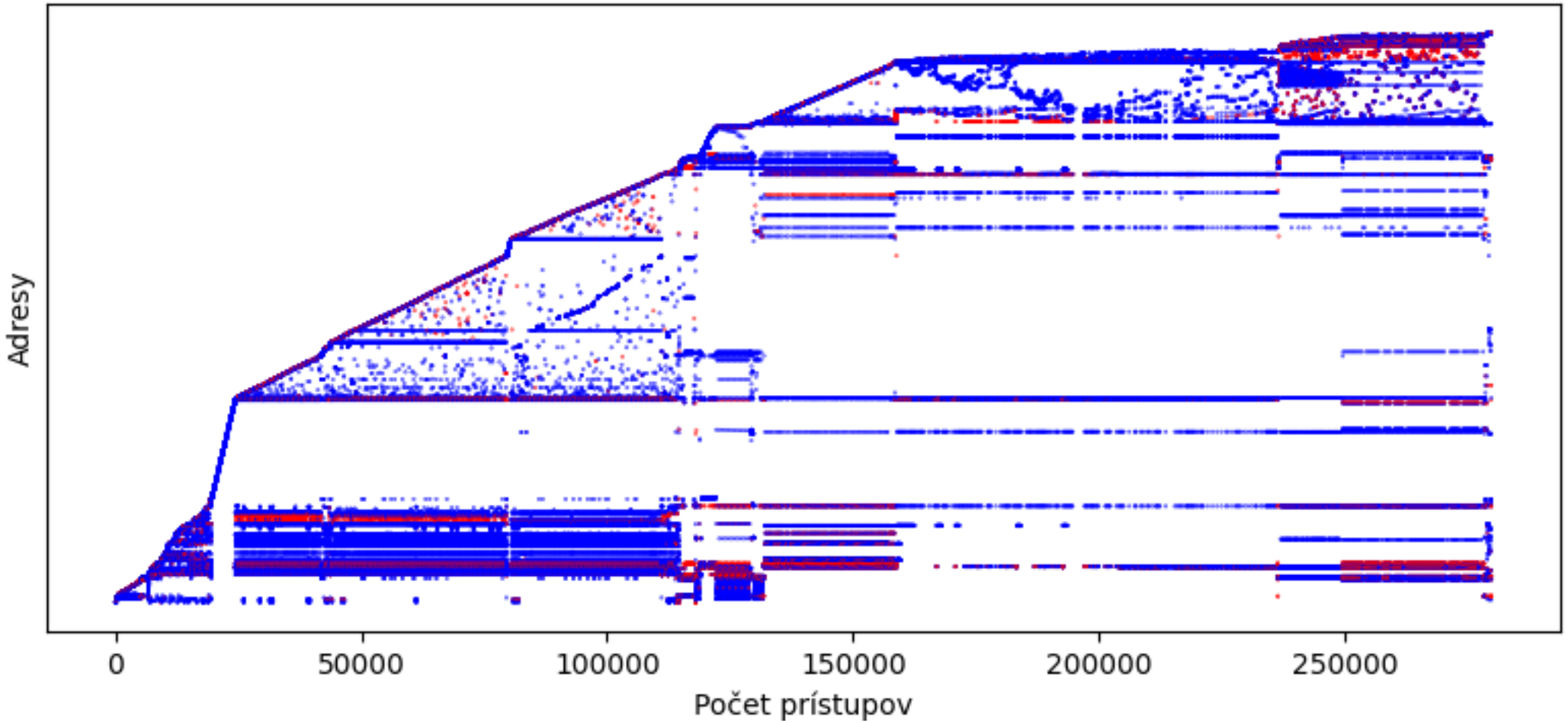
```
0x7f446c455e5f: R 0x7f446c488e70
0x7f446c455e69: R 0x7f446c489000
0x7f446c455e70: W 0x7f446c489a58
0x7f446c455e77: W 0x7f446c489a48
```

```
0xb6d79a38,mov
0xb6d79a3c,bl
0xb6d7a338,ldr
```

Prístupové vzory

- Vyhľadanie prístupových vzorov zadefinovaných Denningom
- Sekvenčný a priestorový
- Vizualizácia v grafe

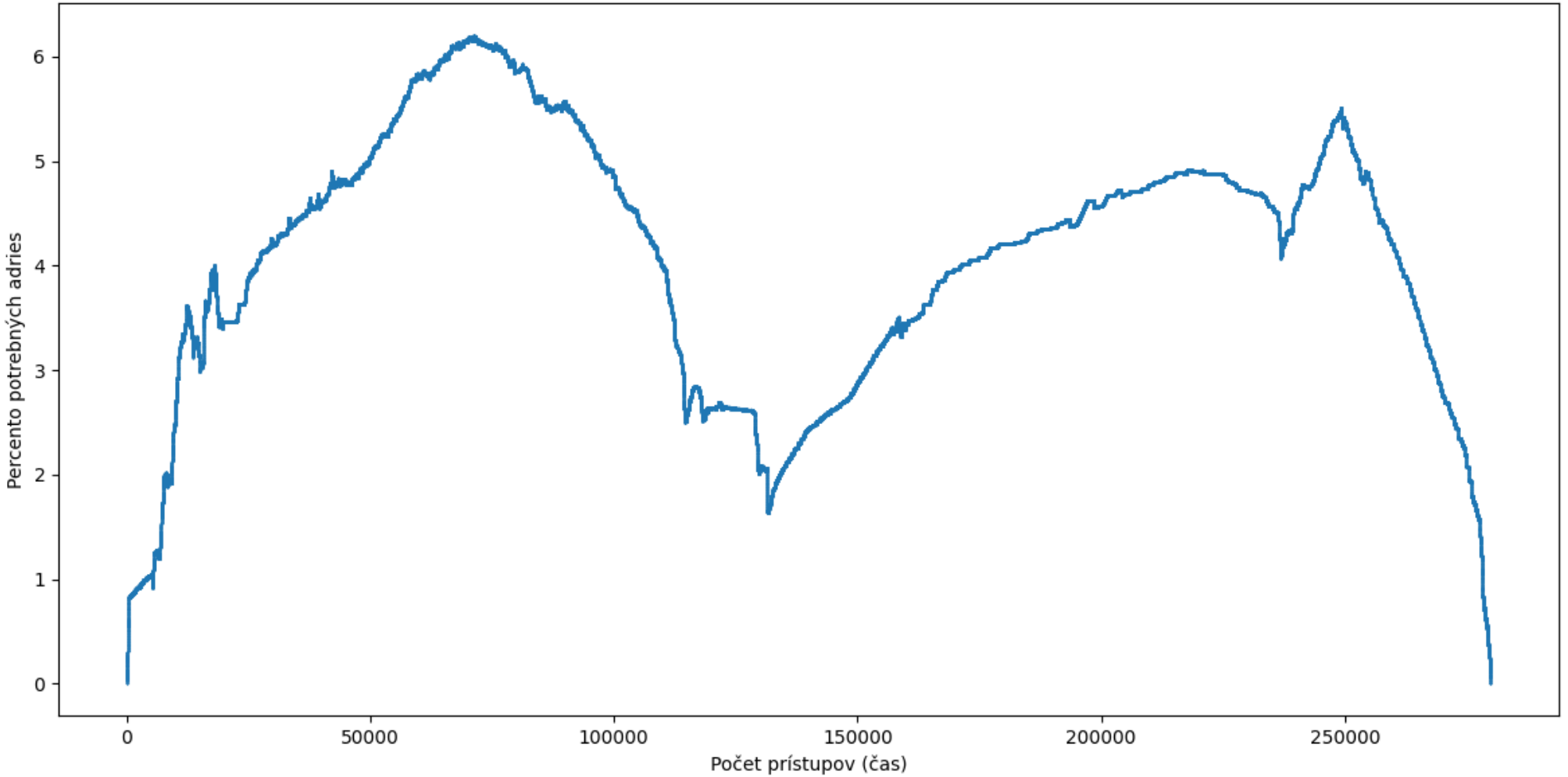
- Väčší podiel vzorov v inštrukčnej pamäti


Adresy
0
50000
100000
150000
200000
250000
Počet prístupov

Odhad minimálneho počtu potrebných adries

- Alokácia registrov
  - Väčší počet premenných/obmedzený počet registrov
- Farbenie grafu – NP-ťažký problém
- Linear scan – prekrývajúce sa intervaly
- Adresám pridelené aktívne intervaly, počet max. aktívnych
- Potrebných adries 4% - 40%, viac v inštrukčnej pamäti

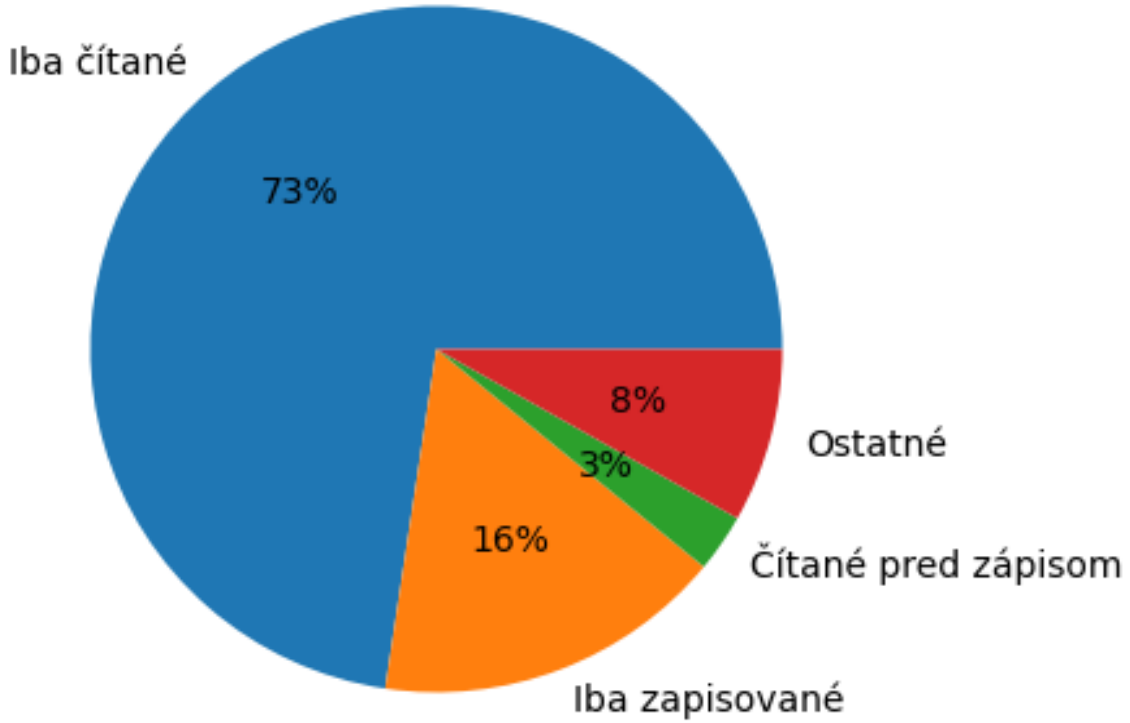
Percento potrebných adries na príkaze ls



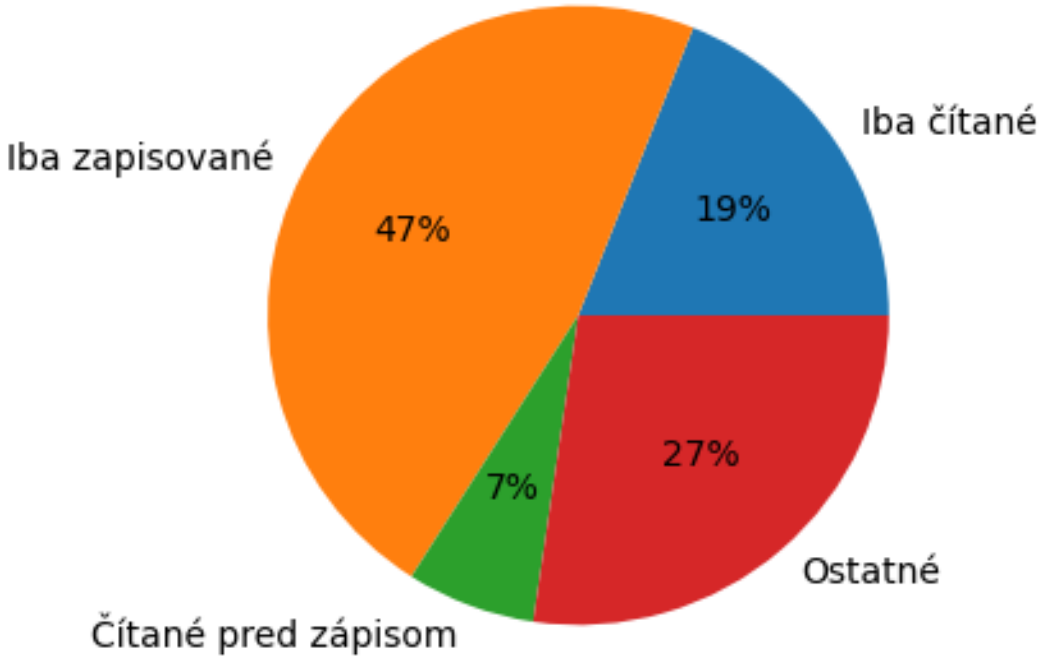
Rozdelenie adries podľa typu

- Adresy, ktoré boli iba čítané, do ktorých bolo iba zapisované, ktoré boli čítané pred zápisom
- Tvorili veľkú časť všetkých adries
- Veľmi odlišné pre rôzne programy
- Komunikácia s operačným systémom, zdieľaná pamäť...

Príkaz *id*



Python *"Hello world!"*



Silno súvislé komponenty

- Vyhľadanie silno súvislých komponentov v grafe pristupovaných adries
  - Rozklad množiny pristupovaných adries

- 1 silno súvislý komponent v dátovej pamäti

- Veľkosti silno súvislých komponentov v inštrukčnej pamäti
- 13905, **37, 31, 10,** 1, 1...

Silno súvislé komponenty

- Prepojenie adries v komponentoch s veľkosťami 37, 31, 10 a */proc/self/maps*

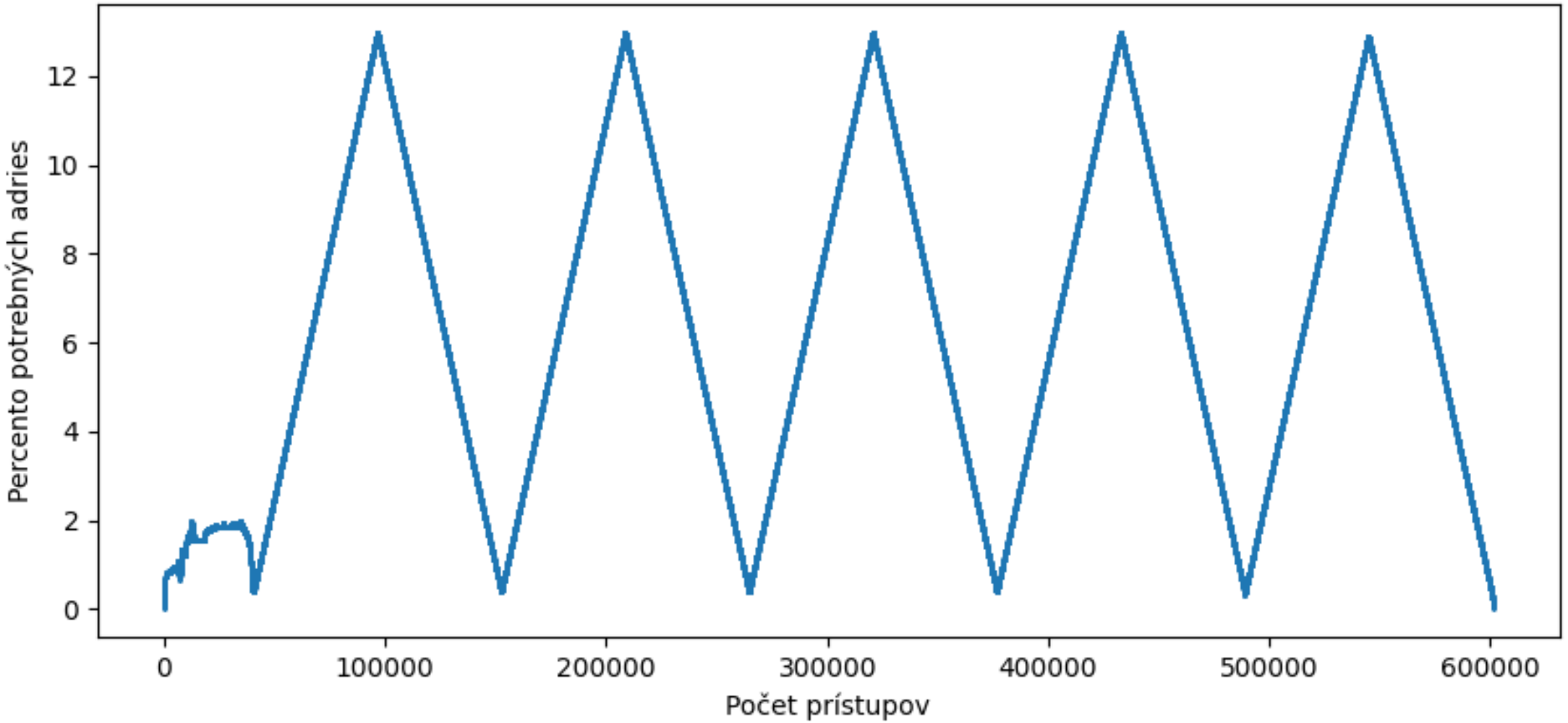
```
7fc531f59000-7fc531f64000 r--p 0002c000 103:06 3279852                   /usr/lib/x86_64-linux-gnu/ld-linux-x86-64.so.2
7fc531f64000-7fc531f65000 r--p 00000000 103:06 1183244                   /usr/share/locale-langpack/en/LC_MESSAGES/gtk30.mo
7fc531f65000-7fc531f67000 r--p 00037000 103:06 3279852                   /usr/lib/x86_64-linux-gnu/ld-linux-x86-64.so.2
7fc531f67000-7fc531f69000 rw-p 00039000 103:06 3279852                   /usr/lib/x86_64-linux-gnu/ld-linux-x86-64.so.2
7ffd7f57d000-7ffd7f59e000 rw-p 00000000 00:00 0                          [stack]
7ffd7f5ee000-7ffd7f5f2000 r--p 00000000 00:00 0                          [vvar]
```

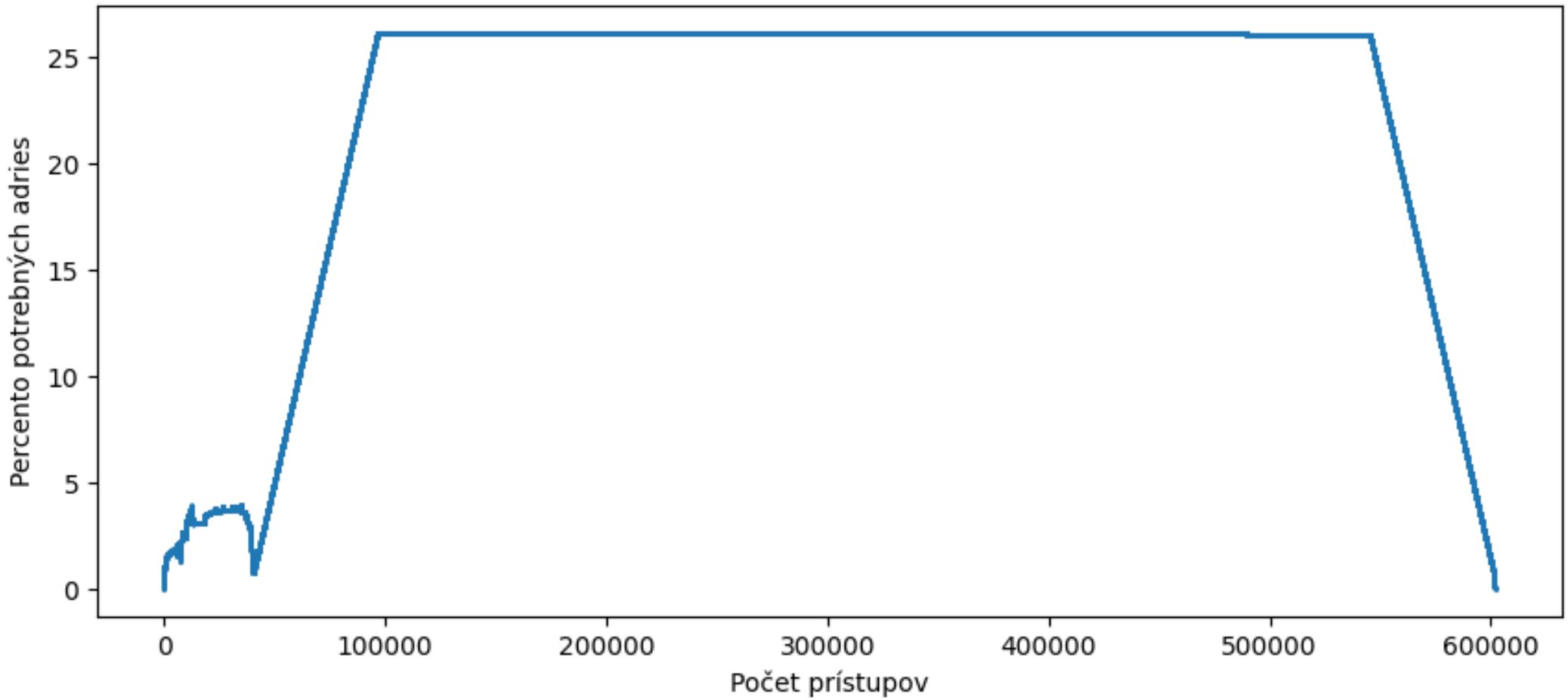
- Dynamický linker
- GCC *-static*
- ARM 35, 28, 7

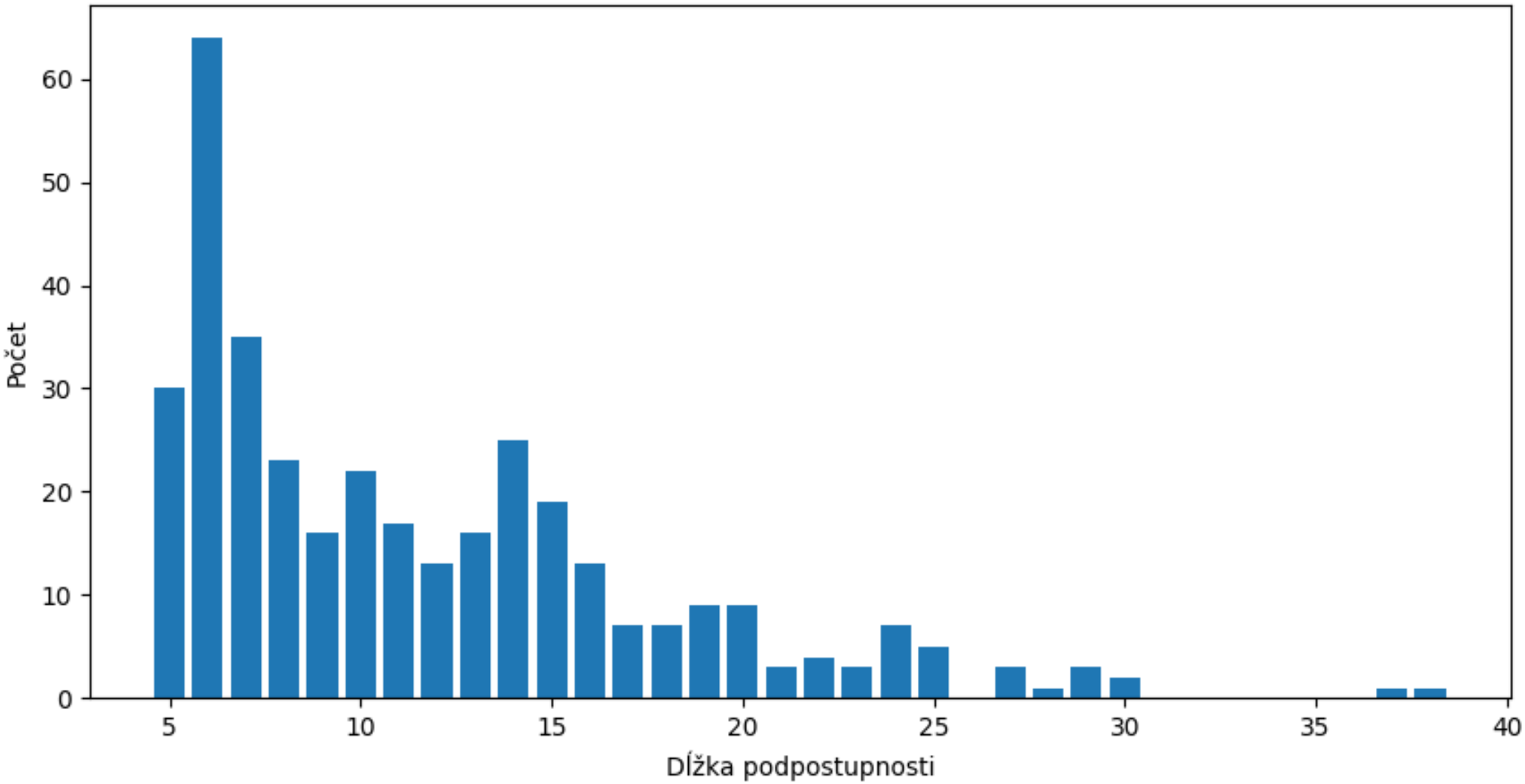
Záver

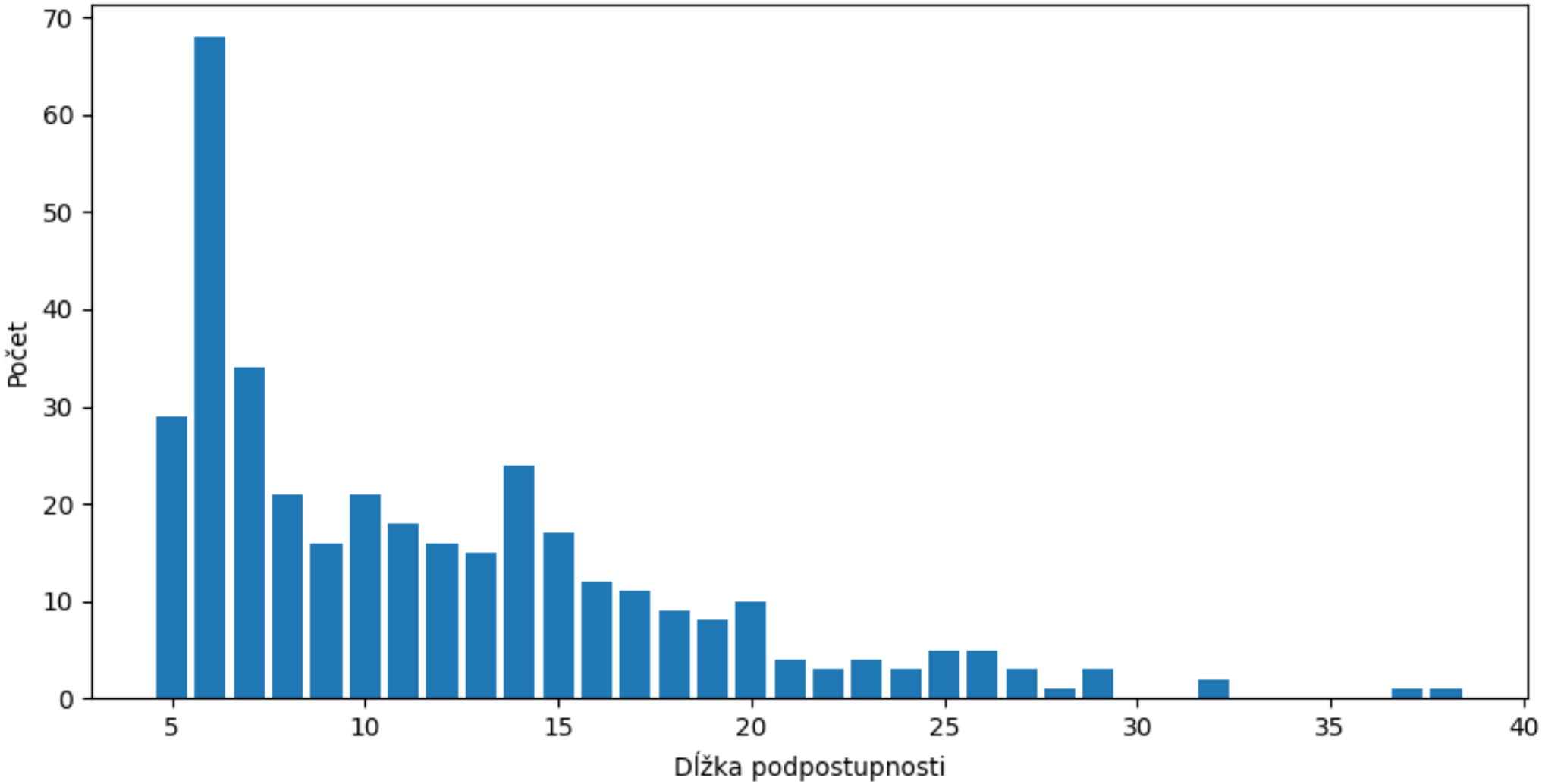
- Vytvorili sme viacero nástrojov v Pythone na analyzovanie prístupov do pamäte
- Nový – grafový prístup sa osvedčil (dynamický linker)
- Využitie vizualizácií a nameraných dát – výučba
- Prepojiť nástroje – hlbšia analýza prístupov do pamäte

Ďakujem za pozornosť


Percento potrebných adries
0
2
4
6
8
10
12
0
100000
200000
300000
400000
500000
600000
Počet prístupov


Percento potrebných adries
25
20
15
10
5
0
0
100000
200000
300000
400000
500000
600000
Počet prístupov


Počet
0
10
20
30
40
50
60
5
10
15
20
25
30
35
40
Dĺžka podpostupnosti


70
60
50
40
30
20
10
0
Počet
5
10
15
20
25
30
35
40
Dĺžka podpostupnosti