

Úvod, obsah kurzu

Paralelní a distribuované systémy, přednáška 1

Tomáš Urbanec

Katedra informatiky PŘF UPOL

25.9.2024

Formality

Organizace a kontakty

Organizace

- Přednáška: Středa 16:45-18:15, 1.127
- Cvičení: Čtvrtek 8:00-9:30, 1.029

Kontakty

- Email: tomas.urbanec@upol.cz
- Telefon: 585 634 713 (pracovna)
- Pracovna: 5.044
- Konzultace ve čtvrtek 10:00-11:00.
- (Nebo po předchozí domluvě.)

Zdroje

- *Foundations of multithreaded, parallel, and distributed programming*, Gregory R. Andrews, Addison-Wesley, 2000 | Knihovna
- *Distributed systems*, Maarten van Steen, Andrew S. Tanenbaum, 4. vyd., distributed-systems.net, 2023 | Knihovna | Online
- *Grokking Concurrency*, Kirill Bobrov, Manning, 2023
- *Principles of concurrent and distributed programming*, Ben-Ari M., Addison-Wesley, 2006 | Knihovna
- Stránky předmětu.
- STAG

Typ kurzu

- Velmi široké téma → projde cokoliv mezi formální logikou a kóděním dnem i nocí.
- Velmi málo času → často nutně přehledově.
- Základní problémy a jejich řešení.
- Zkusíme se udržet okolo prakticky využitelných znalostí.
- Občas zabrousíme hlouběji.
- Zájemci budou mít dostatek zdrojů pro výlet hlouběji vždy.

Zkouška

Zkouška se skládá ze tří částí:

1. Obecný rozhovor o všech tématech kurzu (bez přípravy; přehledově).
2. Klasicky vylosovaná otázka s písemnou přípravou.
3. Rozhovor o jednom vylosovaném článku s písemnou přípravou.

Příčemž:

- Pro splnění zkoušky musí student splnit všechny tři části (v jednom termínu).
- Nejprve si popovídáme v duchu bodu 1. Pokud to bude v pořádku, přistoupíme k losování otázky a článku z bodů 2 a 3.
- Písemná příprava na části 2 a 3 bude max. 30 minut.
- Článek bude během přípravy i rozhovoru k dispozici k nahlédnutí.
- Seznam otázek a článků ke zkoušce bude zveřejněn během semestru.

Úvod do PDS

Několik častých pojmů - znáte je?

- Multitasking (kooperativní, preemptivní)
- Sekvenční výpočet
- Konkurentní výpočet
- Paralelní výpočet
- Distribuovaný výpočet
- Asynchronní výpočet

Co budeme řešit?

Paralelní systém

- Využití více jader/CPU/GPU
- Sdílená paměť
- Synchronizace

Distribuovaný systém

- Využití více strojů/systémů
- Synchronizace po síti
- Replikace, vzájemné vyloučení, synchronizace času, globální stav, ...

Proč to chceme řešit?

- Rychlejší zpracování dat (P i D)
- Zpracování více dat (P i D)
- Více uživatelů (P i D)
- Decentralizace (D)
- Dostupnost (spolehlivost) systému (D)
- Dostupnost (geografická) systému (D)

Příklady

Dnes všude.

- Weby
- Hry
- Sociální sítě
- Kryptoměny
- Superpočítače
- Cloud
- Databáze
- Mobilní síť
- Internet (!)

Problémy

- Správnost
- Synchronizace, konzistence
- Debugování
- Intuice

Jak na to?

- Vlákna
- Synchronizace
- Procesy
- Komunikace
- Sítě
- Abstrakce

Synchronizace

Paralelní

- Sdílená paměť
- Atomicita
- Synchronizační primitiva a bariéry
- Časté vzory → klasické problémy
- Paralelní algoritmy

Distribuované

- Architektura
- Komunikace, koordinace, konzistence, replikace, tolerance chyb, ...
- Časté vzory → klasické problémy
- Distribuované algoritmy

Příklady

Násobení matic

Motivační příklad

Na tabuli:

- Sekvenčně
- Paralelně
- Distribuovaně

Zápis do ...

Motivační příklad

... souboru, databáze, socketu, ...

Na tabuli (DB):

- Sekvenčně
- Paralelně
- Distribuovaně

Reálný svět

Několik dalších pojmů

- CPU - méně jader, plná instrukční sada
- (GP)GPU - mnoho jader, omezená instrukční sada (vektorové operace)
- Cluster - více strojů v síti; typicky u sebe; těsně propojené; abstrakce za API
- Superpočítač - cluster, typicky s velmi rychlými propoji
- Grid computing - více strojů v síti; klidně daleko; volnější než cluster
- Cloud - transparentní přístup k ...
 - hardware (i virtualizovaný) → IaaS (AWS, Azure, ...),
 - software → SaaS (Google Apps, Office365, ...),
 - platformě (včetně vývoje) → PaaS (Salesforce, Google, ...)

Flynnova taxonomie

- Michael J. Flynn, 1966 a 1972
- Význam: **S**ingle, **M**ultiple, **I**nstruction, **D**ata
- **SISD** - von Neumann
- **SIMD** - vektorové instrukce, grafické karty
- **MISD** - neuronové sítě, systolické pole
- **MIMD** - dnes téměř vše

Dnes už nestačí (MIMD).

Existují nastavby a různá jemnější dělení.

- SMP (symetrický multiprocesor) - jako mutlijádro ale o úroveň výše.
- DSM (distribuovaná sdílená paměť) - (po síti) sdílená (virtuální) paměť
- SPMD (**P**rogram) - jako SIMD ale ne synchronizované
- ...

Kdy paralelizovat/distribuuovat?

Let me put it this way: During the past 50 years, I've written well over a thousand programs, many of which have substantial size. I can't think of even five of those programs that would have been enhanced noticeably by parallelism or multithreading.

Donald E. Knuth [1]

So I decided long ago to stick to what I know best. Other people understand parallel machines much better than I do; programmers should listen to them, not me, for guidance on how to deal with simultaneity.

Donald E. Knuth [1]

A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable.

Leslie Lamport [?]

We know of no area in computer science or mathematics in which informal reasoning is more likely to lead to errors than in the study of this type of [distributed] algorithm.

Leslie Lamport et al. [2]

Pokud nemusíte, tak to nedělejte.

Kdy paralelizovat/distribuuovat?

- Výkon
- Spolehlivost
- Fyzické a fyzikální důvody
- Bezpečnost
- Dostupnost
- ...

Ale dnes (relativně) často musíte.

Škálování

Zvyšování výkonu

- Vertikální (up) škálování
 - Lepší procesor,
 - rychlejší disk,
 - více paměti,
 - ...
- Horizontální (out) škálování
 - více strojů.

Příklad

Škálování a složitost systému

- Postupný vývoj úžasné webové služby.

Škálování a údržba

- Jak často očekávat potíže?
- Při jednom stroji?
- Při 1000 strojích?

Zrychlení výpočtu

Více pohledů:

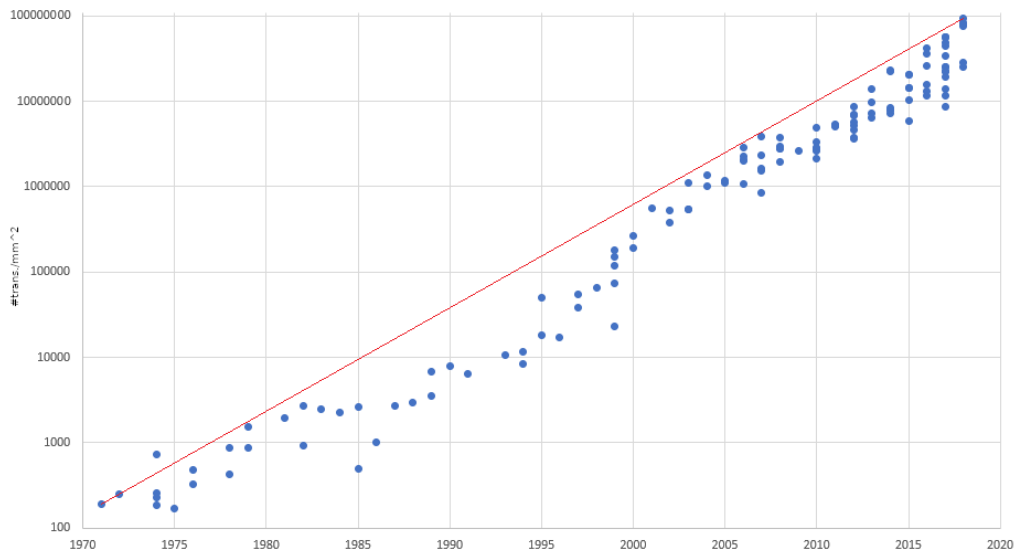
- Moorův zákon (1965) - výkon (počet tranzistorů) CPU se každých 18 měsíců zdvojnásobí.
- Amdahlův zákon (1967) - použijeme více procesorů pro řešení stejného problému
- Gustafson-Barsisův zákon (1988) - když máme více zdrojů, můžeme problém řešit lépe/detailněji

N - počet procesorů,

$F = \frac{T_s}{T_s + T_p}$, kde T_s je neparalelizovatelná část a T_p je paralelizovatelná část programu.

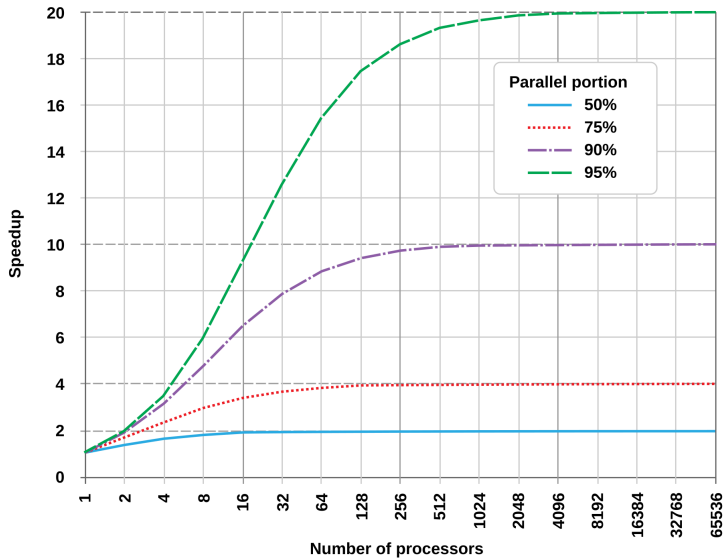
Amdahl	Gustafson-Barsis
zrychlení = $\frac{1}{F + \frac{1-F}{N}}$	zrychlení = $F + (1 - F) \cdot N$

Moorův zákon



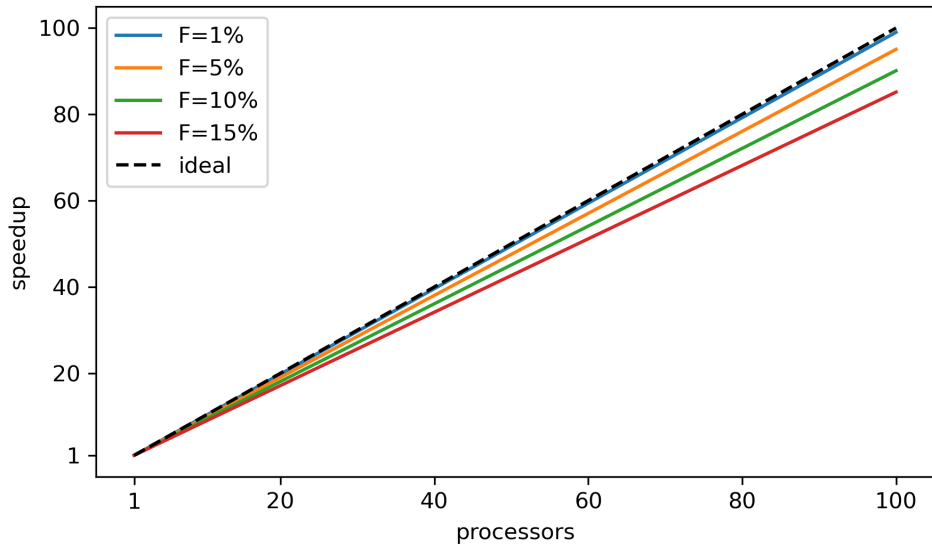
Obrázek: Moorův zákon. Zdroj grafu [3]

Amdahlův zákon



Obrázek: Amdahlův zákon. Zdroj grafu [4]

Gustafson-Barsisův zákon



Obrázek: Gustafson-Barsisův zákon. Zdroj grafu [5]

Zrychlení výpočtu

- Moorův zákon a fyzikální limity.
- Amdahlův zákon je limitovaný velikostí problému, reálně lepší výsledky.
- Gustafson-Barsis: Paralelní část systému škáluje se zdroji, sériová část se nemění
→ lepší využití výkonu.

Průřez kurzem

Paralelní systémy






Zhruba 3-4 přednášky.

- Chyby souběhu
- Synchronizace sdílenou pamětí
- Problém kritické sekce
- Synchronizační primitiva
- Bariéry a další vzory
- Klasické problémy
- Návrh paralelního systému (Foster)

Distribuované systémy

Zhruba 7-8 přednášek.

- Architektura
 - Client-server
 - P2P
 - Middleware
 - Virtualizace
- Komunikace
 - Zasílání zpráv
 - RPC
 - Multicast
- Koordinace
 - Čas
 - Vzájemné vyloučení
 - Volba leadera
- Odolnost chybám
- Konzistence, transakce
- Bezpečnost
- Blockchain

-  Andrew Binstock. Rozhovor s Donaldem E. Knuthem. Dostupné z <http://www.informit.com/articles/article.aspx?p=1193856>. Citováno 24.9.2024.
-  Leslie Lamport, Robert Shostak, and Marshall Pease. 1982. The Byzantine Generals Problem. ACM Trans. Program. Lang. Syst. 4, 3 (July 1982), 382–401. <https://doi.org/10.1145/357172.357176>
-  Graf k Moorovu zákonu. <https://medium.com/predict/moores-law-is-alive-and-well-eaa49a450188>. Citováno 24.9.2024.
-  Graf k Amdahlovu zákonu. https://en.wikipedia.org/wiki/Amdahl%27s_law#cite_note-2. Citováno 24.9.2024.
-  Graf k Gustafosn-Barisovu zákonu. https://hpc-wiki.info/hpc/Benchmarking_%26_Scaling_Tutorial/Introduction. Citováno 24.9.2024.

Changelog

- 11.10.2024
 - styl úvodního slidu,
 - nastavení beameru,
 - changelog,
 - překlepy.