

# Vypracované otázky k MSZ pro rok 2022

Specializace NNET

## **Specializace Počítačové sítě – NNET**

1. Architektura superskalárních procesorů a algoritmy zpracování instrukcí mimo pořadí, predikce skoků.
2. Paměťová konzistence a předbíhání operací čtení a zápisu, podpora virtuálního adresového prostoru.
3. Datový paralelismus SIMD, HW implementace a SW podpora.
4. Architektury se sdílenou pamětí UMA a NUMA, zajištění lokality dat.
5. Problém koherence pamětí cache na systémech se sdílenou pamětí, protokol MSI.
6. Paralelní zpracování v OpenMP: Smyčky, sekce a tasky a synchronizační prostředky.
7. Pravděpodobnost, podmíněná pravděpodobnost, nezávislost.
8. Náhodná proměnná, typy náhodné proměnná, funkční a číselné charakteristiky, významná rozdělení pravděpodobnosti.
9. Bodové a intervalové odhady parametrů, testování hypotéz o parametrech.
10. Vícevýběrové testy, testy o rozdělení, testy dobré shody.
11. Regresní analýza.
12. Markovské řetězce a základní techniky pro jejich analýzu.
13. Randomizované algoritmy (Monte Carlo a Las Vegas algoritmy).
14. Problém generalizace strojového učení a přístup k jeho řešení (trénovací, validační a testovací sada, regularizace, předtrénování, multi-task learning, augmentace dat, dropout, ...)
15. Generativní modely a diskriminativní přístup ke klasifikaci (gaussovský klasifikátor, logistická regrese, ...)
16. Neuronové sítě a jejich trénování (metoda gradientního sestupu, účelová (loss) funkce, výpočetní graf, aktivační funkce, zápis pomocí maticového násobení, ...)
17. Neuronové sítě pro strukturovaná data (konvoluční a rekurentní sítě, motivace, základní vlastnosti, použití)
18. Prohledávání stavového prostoru (informované a neinformované metody, lokální prohledávání, prohledávání v nejistém prostředí, hraní her, CSP úlohy)
19. Klasifikace formálních jazyků (Chomského hierarchie), vlastnosti formálních jazyků a jejich rozhodnutelnost.
20. Konečné automaty (jazyky přijímané KA, varianty KA, minimalizace KA, Mihill-Nerodova věta).
21. Regulární množiny, regulární výrazy a rovnice nad regulárními výrazy.
22. Zásobníkové automaty (jazyky přijímané ZA, varianty ZA).
23. Turingovy stroje (jazyky přijímané TS, varianty TS, lineárně omezené automaty, vyčíslitelné funkce).
24. Nerozhodnutelnost (problém zastavení TS, princip diagonalizace a redukce, Postův korespondenční problém).
25. Časová a paměťová složitost (třídy složitosti, úplnost, SAT problém).
26. Postrelační a rozšířené relační databáze (objektový a objektově relační databázový model – struktura a operace; podpora práce s XML a JSON dokumenty v databázích).
27. NoSQL databáze (porovnání relačních a NoSQL; CAP věta a ACID/BASE principy; typy NoSQL databází; dotazování v NoSQL databázích; agregace dat pomocí Map-Reduce a agregační pipeline).
28. Získávání znalostí z dat (pojem znalost; typické zdroje dat; základní úlohy získávání znalostí; analytické projekty a proces získávání znalostí z dat).

29. Porozumění datům (důvod a cíl; popisné charakteristiky dat a vizualizační techniky; korelační analýza).
30. Prostorové DB (problematika mapování prostoru, ukládání, indexace; využití).
31. Indexace (nejen) v prostorových DB (kD-Tree a Grid File (a jejich varianty), R-Tree).
32. Lambda kalkul (definice všech pojmu, operací...).
33. Práce v lambda kalkulu (demonstrace reprezentace čísel a pravdivostních hodnot a operací nad nimi).
34. Haskell – lazy evaluation (typy v jazyce včetně akcí, uživatelské typy, význam typových tříd, demonstrace lazy evaluation).
35. Prolog – způsob vyhodnocení (základní princip, unifikace, chování vestavěných predikátů, operátor řezu – vhodné a nevhodné užití).
36. Prolog – změna DB/programu za běhu (demonstrace na prohledávání stavového prostoru, práce se seznamy).
37. Model PRAM, suma prefixů a její aplikace.
38. Distribuované a paralelní algoritmy – algoritmy nad seznamy, stromy a grafy.
39. Interakce mezi procesy a typické problémy paralelismu (synchronizační a komunikační mechanismy).
40. Distribuované a paralelní algoritmy – předávání zpráv a knihovny pro paralelní zpracování (MPI).
41. Distribuovaný broadcast, synchronizace v distribuovaných systémech.
42. Klasifikace a vlastnosti paralelních a distribuovaných architektur, základní typy jejich topologií.
43. Distribuované a paralelní algoritmy – algoritmy řazení, select, algoritmy vyhledávání.
44. Bezdrátové lokální sítě (Wifi, Bluetooth).
45. Hledání minimální kostry obyčejného grafu (pojmy, stromy a kostry, Kruskalův algoritmus, Primův algoritmus).
46. Hledání nejkratších cest ze zdrojového uzlu do všech ostatních uzel grafu (Bellman-Fordův algoritmus, Dijkstrův algoritmus).
47. Klasifikace algoritmů volby koordinátora, algoritmus Bully a jeho složitost.
48. Podmínky konsistentního globálního stavu distribuovaného systému.
49. Principy distribuovaného zpracování MapReduce, průběh a jednotlivé operace distribuovaného výpočtu pomocí MapReduce, jeho implementace v Apache Hadoop a Apache Spark.
50. Symetrická kryptografie. Vlastnosti, vlastnosti bezpečného algoritmu, délka klíče, útok silou, příklady symetrických algoritmů, Feistelovy šifry, DES, režimy činnosti, proudové šifry.
51. Asymetrická kryptografie, vlastnosti, způsoby použití, poskytované bezpečnostní funkce, elektronický podpis a jeho vlastnosti, hybridní kryptografie, algoritmus RSA, generování klíčů, šifrování, dešifrování.
52. Hašovací funkce, klíčovaný haš a MAC a jejich použití a vlastnosti.
53. Správa klíčů v asymetrické kryptografii (certifikáty X.509).
54. Základní architektury přepínačů, algoritmy pro plánování, řešení blokování, vícestupňové přepínací sítě.
55. Základní funkce směrovače, zpracování paketů ve směrovači, typy přepínání a architektur.
56. Metody pro výpočet směrování v sítích (Bellman-Ford, Dijkstra, Path vector, DUAL).
57. Řízení toku dat (flow-control) a prevence zahlcení (congestion-control) na transportní vrstvě (MP-TCP, QUIC, SCTP, DCCP).
58. Metody detekce síťových incidentů (signatury, statistické metody) a nástroje (IDS/IPS).
59. Sítě Peer-to-Peer: vlastnosti, chování, způsoby směrování. Strukturované a nestrukturované sítě.

60. Události v JavaScriptu (smyčka událostí, asynchronní programování, klientské události, obsluha událostí)

61. Přenos a distribuce webových dat (URI, protokol HTTP, proxy HTTP, CDN, XHR)

62. Bezpečnost webových aplikací (SOP, XSS, CSRF, bezpečnostní hlavičky HTTP)

# Obsah

|    |   |     |
|----|---|-----|
| 1  | TIN – Klasifikace formálních jazyků (Chomského hierarchie), vlastnosti formálních jazyků a jejich rozhodnutelnost.  | 5   |
| 2  | TIN – Konečné automaty (jazyky přijímané KA, varianty KA, minimalizace KA, Mihill-Nerodova věta).   | 19  |
| 3  | TIN – Regulární množiny, regulární výrazy a rovnice nad regulárními výrazy.   | 27  |
| 4  | TIN – Zásobníkové automaty (jazyky přijímané ZA, varianty ZA).  | 30  |
| 5  | TIN – Turingovy stroje (jazyky přijímané TS, varianty TS, lineárně omezené automaty, vyčíslitelné funkce).  | 34  |
| 6  | TIN – Nerozhodnutelnost (problém zastavení TS, princip diagonalizace a redukce, Postův korespondenční problém).   | 43  |
| 7  | TIN – Časová a paměťová složitost (třídy složitosti, úplnost, SAT problém).   | 55  |
| 8  | GAL – Hledání minimální kostry obyčejného grafu (pojmy, stromy a kostry, Kruskalův algoritmus, Primův algoritmus).  | 56  |
| 9  | GAL – Hledání nejkratších cest ze zdrojového uzlu do všech ostatních uzel grafu (Bellman-Fordův algoritmus, Dijkstrův algoritmus).  | 66  |
| 10 | PDI – Klasifikace algoritmů volby koordinátora, algoritmus Bully a jeho složitost.  | 75  |
| 11 | PDI – Podmínky konsistentního globálního stavu distribuovaného systému.   | 81  |
| 12 | PDI – Principy distribuovaného zpracování MapReduce, průběh a jednotlivé operace distribuovaného výpočtu pomocí MapReduce, jeho implementace v Apache Hadoop a Apache Spark.  | 84  |
| 13 | KRY – Symetrická kryptografie. Vlastnosti, vlastnosti bezpečného algoritmu, délka klíče, útok silou, příklady symetrických algoritmů, Feistelovy šifry, DES, režimy činnosti, proudové šifry.                         | 91  |
| 14 | KRY – Asymetrická kryptografie, vlastnosti, způsoby použití, poskytované bezpečnostní funkce, elektronický podpis a jeho vlastnosti, hybridní kryptografie, algoritmus RSA, generování klíčů, šifrování, dešifrování. | 105 |
| 15 | KRY – Hašovací funkce, klíčovaný haš a MAC a jejich použití a vlastnosti.   | 110 |

|  |     |
|--|-----|
| 16 KRY – Správa klíčů v asymetrické kryptografii (certifikáty X.509).  | 117 |
| 17 PDS – Prerekvizity k ostatním otázkám.  | 123 |
| 18 PDS – Základní architektury přepínačů, algoritmy pro plánování, řešení blokování, vícestupňové přepínací sítě.                  | 127 |
| 19 PDS – Základní funkce směrovače, zpracování paketů ve směrovači, typy přepínání a architektur.                                  | 141 |
| 20 PDS – Metody pro výpočet směrování v sítích (Bellman-Ford, Dijkstra, Path vector, DUAL).  | 157 |
| 21 PDS – Řízení toku dat (flow-control) a prevence zahlcení (congestion-control) na transportní vrstvě (MP-TCP, QUIC, SCTP, DCCP). | 163 |
| 22 PDS – Metody detekce sítových incidentů (signatury, statistické metody) a nástroje (IDS/IPS).                                   | 178 |
| 23 PDS – Sítě Peer-to-Peer: vlastnosti, chování, způsoby směrování. Strukturované a nestrukturované sítě.                          | 185 |

# Kapitola 1

## TIN – Klasifikace formálních jazyků (Chomského hierarchie), vlastnosti formálních jazyků a jejich rozhodnutelnost.

### 1.1 Zdroje

- tin\_2021\_merged.pdf
- TIN\_2020-09-22.mp4
- TIN\_2020-09-29.mp4
- TIN\_2020-10-06.mp4
- TIN\_2020-10-13.mp4
- <https://petrzemek.net/publications>

### 1.2 Úvod a kontext

- Symbol, znak
- Abeceda – Konečná množina symbolů.
- Řetězec
- Konkatenace řetězců
- Reverze řetězce
- Podřetězec
- Délka řetězce
- Formální jazyk
- Doplňek jazyka
- Konkatenace jazyků
- Mocnina jazyka
- Iterace jazyka

### 1.3 Gramatiky

**Gramatika** Gramatika slouží k formální specifikaci jazyků, zejména pak nekonečných jazyků. Gramatika je čtverice  $G = (N, \Sigma, P, S)$ , kde

- $N$  je konečná množina neterminálů (pomocné syntaktické celky);
- $\Sigma$  je konečná množina terminálů (symbolů),
  - $N \cap \Sigma = \emptyset$ ;
- $P$  je konečná množina pravidel,
  - $P \subseteq (N \cup \Sigma)^* N (N \cup \Sigma)^* \times (N \cup \Sigma)^*$ ;
  - prvek  $(\alpha, \beta) \in P$  zapisujeme jako  $\alpha \rightarrow \beta$ ;
- $S$  je výchozí neterminál,
  - $S \in N$ .

**Přímá derivace** Necht'  $G = (N, \Sigma, P, S)$  je gramatika a necht'  $\lambda, \mu \in (N \cup \Sigma)^*$ . Mezi řetězci  $\lambda$  a  $\mu$  platí binární relace  $\lambda \Rightarrow \mu$ , nazývaná přímá derivace, pokud můžeme vyjádřit:

$$\lambda = \gamma \alpha \delta$$

$$\mu = \gamma \beta \delta$$

kde  $\gamma, \delta \in (N \cup \Sigma)^*$  a  $\alpha \rightarrow \beta \in P$ . Říkáme, že řetězec  $\mu$  lze přímo derivovat z řetězce  $\lambda$  v gramatice  $G$ .

**Derivace** Necht'  $G = (N, \Sigma, P, S)$  je gramatika a necht'  $\lambda, \mu \in (N \cup \Sigma)^*$ . Mezi řetězci  $\lambda$  a  $\mu$  platí binární relace  $\lambda \Rightarrow^+ \mu$ , nazývaná derivace, jestliže existuje posloupnost přímých derivací  $v_{i-1} \Rightarrow v_i, \quad i = 1, \dots, n, \quad n \geq 1$ , taková, že platí:

$$\lambda = v_0 \Rightarrow v_1 \Rightarrow \dots \Rightarrow v_{n-1} \Rightarrow v_n = \mu$$

. Tuto posloupnost nazýváme derivací délky  $n$ . Říkáme, že řetězec  $\mu$  lze derivovat z řetězce  $\lambda$  v gramatice  $G$ . Jestliže v gramatice  $G$  platí pro řetězce  $\lambda$  a  $\mu$  relace  $\lambda \Rightarrow^+ \mu$  a nebo identita  $\lambda = \mu$ , pak píšeme  $\lambda \Rightarrow^* \mu$ .

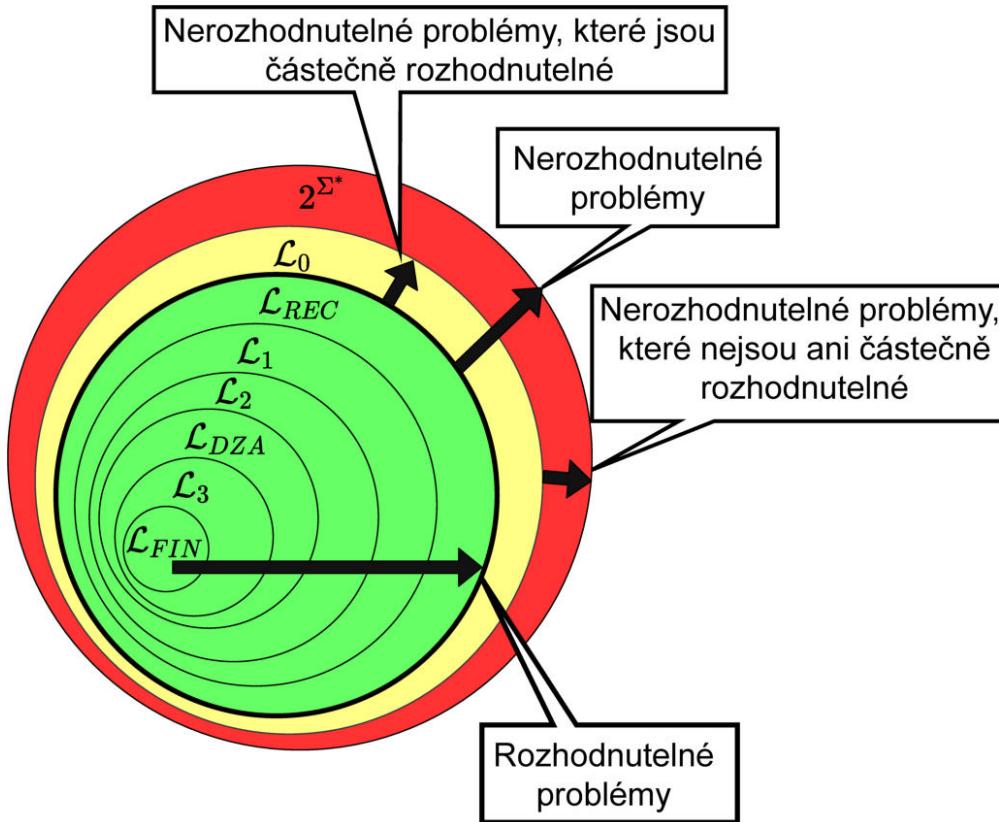
**Větná forma, věta, jazyk generovaný gramatikou** Necht'  $G = (N, \Sigma, P, S)$  je gramatika. Řetězec  $\alpha \in (N \cup \Sigma)^*$  nazýváme větnou formou, jestliže platí  $S \Rightarrow^* \alpha$ , tj. řetězec  $\alpha$  je generovatelný z výchozího symbolu  $S$ . Větná forma, které obsahuje pouze terminální symboly, se nazývá věta. Jazyk  $L(G)$  generovaný gramatikou  $G$ , je definován množinou všech vět

$$L(G) = \{w \mid w \in \Sigma^* \wedge S \Rightarrow^+ w\}$$

### 1.4 Chomského hierarchie

Chomského hierarchie je hierarchie tříd formálních gramatik generujících formální jazyky. Gramatiky jsou děleny dle tvaru jejich pravidel do 4 kategorií. Každé kategorie gramatik odpovídá kategorie formálních jazyků, které jsou gramatikami generovány –  $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ .

Platí:  $2^{\Sigma^*} \supset \mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \mathcal{L}_3$ . Existují i jazyky mimo třídu  $\mathcal{L}_0$ , které není možné gramatikami vyjádřit. Chomského hierarchie vymezuje popisnou a rozhodovací sílu dané třídy.



Obrázek 1.1: Chomského hierarchie doplněna o další třídy a podtřídy jazyků. Jednotlivé třídy jsou vysvětleny dále. Teorii nerozhodnutelnosti je věnována jiná otázka.

### Třída jazyků $2^{\Sigma^*} - \mathcal{L}_0$

- Jazyky v této třídě není možné vyjádřit žádnou gramatikou.

### Typ 0 (obecné / neomezené / rekurzivně vyčíslitelné gramatiky)

- Pravidla v nejobecnějším tvaru:
  - $\alpha \rightarrow \beta, \quad \alpha \in (N \cup \Sigma)^* N (N \cup \Sigma)^*, \quad \beta \in (N \cup \Sigma)^*$
- Jazyky  $\mathcal{L}_0$  jsou přijímány turingovými stroji (TS) s nekonečnou páskou.
  - Varianty: deterministický TS, nedeterministický TS a vícepáskový TS disponují stejnou vyjadřovací silou (jsou mezi sebou převoditelné).
- Mezi třídami jazyků  $\mathcal{L}_0$  a  $\mathcal{L}_1$  existuje třída tzv. rekurzivních jazyků ( $\mathcal{L}_{Rec}$ , Recursive Languages).
  - Platí  $\mathcal{L}_0 \supset \mathcal{L}_{Rec} \supset \mathcal{L}_1$ .
  - Rekurzivně vyčíslitelné jazyky:

- \* Nechť  $M$  je turingův stroj, který přijímá nějaký jazyk z  $\mathcal{L}_0$ . Pokud  $M$  dostane na vstup  $w \in L(M)$ , tak pro něj zastaví a přijme ho. Avšak, pokud  $M$  dostane na vstup  $w \notin L(M)$ , tak bud' zastaví a odmítne ho a nebo se zacyklí (nezastaví).
- Rekurzivní jazyky:
  - \* Nechť  $M$  je turingův stroj, který přijímá nějaký jazyk z  $\mathcal{L}_{Rec}$ . Pokud  $M$  dostane na vstup  $w \in L(M)$ , tak pro něj zastaví a přijme ho. Avšak, pokud  $M$  dostane na vstup  $w \notin L(M)$ , tak zastaví a odmítne ho. Turingův stroj, který zastaví pro jakýkoliv vstup se nazývá úplný turingův stroj.

## Typ 1 (kontextové gramatiky)

- Pravidla ve tvaru:
  - $\alpha A\beta \rightarrow \alpha\gamma\beta$ ,  $A \in N$ ,  $\alpha, \beta \in (N \cup \Sigma)^*$ ,  $\gamma \in (N \cup \Sigma)^+$
  - A nebo  $S \rightarrow \epsilon$ , pokud se  $S$  nevyskytuje na pravé straně žádného pravidla (neobsahuje tzv. epsilon pravidla).
  - Z tvaru pravidel vyplývá, že větné formy se v průběhu derivace nemohou zkraťovat, ledaže se derivuje  $S \Rightarrow \epsilon$ , ale v takovém případě se  $S$  nemůže vyskytovat na pravé straně žádného pravidla.
  - Formálně: pokud  $\alpha \Rightarrow \beta$ , pak  $|\alpha| \leq |\beta|$  (s vyjímkou  $S \Rightarrow \epsilon$ ).
- Jazyky  $\mathcal{L}_1$  jsou přijímány nedeterministickým lineárně omezeným automatem (LOA).
  - Nedeterministický LOA je NTS s lineárně omezenou páskou v závislosti na délce vstupního řetězce.
  - Není známo, zda deterministický LOA disponuje stejnou vyjadřovací silou jako nedeterministický LOA.
- Příklad jazyků:
  - $\{a^n b^n c^n \mid n \geq 10\} \in \mathcal{L}_1 \setminus \mathcal{L}_2$

## Typ 2 (bezkontextové gramatiky)

- Pravidla ve tvaru:
  - $A \rightarrow \alpha$ ,  $A \in N$ ,  $\alpha \in (N \cup \Sigma)^*$
  - Mohou obsahovat i tzv. epsilon pravidla (narozdíl od kontextových).
  - Jak je to možné, když  $\mathcal{L}_1 \supset \mathcal{L}_2$ ?
  - Bezkontextové gramatiky s epsilon pravidly jsou převoditelné na kontextové gramatiky bez epsilon pravidel.
- Jazyky  $\mathcal{L}_2$  jsou přijímány nedeterministickým zásobníkovým automatem (ZA).
  - Varianty: nedeterministický ZA a rozšířený nedeterministický ZA disponují stejnou vyjadřovací silou.
- Mezi třídami jazyků  $\mathcal{L}_2$  a  $\mathcal{L}_3$  existuje třída tzv. deterministických bezkontextových jazyků ( $\mathcal{L}_{DCF}$ , *Deterministic Context Free Languages*).
  - Platí  $\mathcal{L}_2 \supset \mathcal{L}_{DCF} \supset \mathcal{L}_3$ .

- Jsou přijímány deterministickým ZA a deterministickým rozšířeným ZA. [[todo: ověřit]]

- Příklad jazyků:

- $\{a^n b^n \mid n \geq 1\} \in \mathcal{L}_{DCF} \setminus \mathcal{L}_3$
- $\{w \mid w \in \{a, b\}^* \wedge \#_a(w) = \#_b(w)\} \in \mathcal{L}_{DCF} \setminus \mathcal{L}_3$
- $\{ww^R \mid w \in \{a, b\}^+\} \in \mathcal{L}_2 \setminus \mathcal{L}_{DCF}$
- $\{a^i b^j c^k \mid i = j \vee j = k\} \in \mathcal{L}_2 \setminus \mathcal{L}_{DCF}$

### Typ 3 (regulární gramatiky)

- Pravidla ve tvaru:
  - Pravolineární:  $A \rightarrow wB \mid w, \quad A, B \in N, \quad w \in \Sigma^*$
  - Levolineární:  $A \rightarrow Bw \mid w, \quad A, B \in N, \quad w \in \Sigma^*$
  - Pravoregulární:  $A \rightarrow aB \mid a, \quad A, B \in N, \quad a \in \Sigma$
  - Levoregulární:  $A \rightarrow Ba \mid a, \quad A, B \in N, \quad a \in \Sigma$
- Gramatiky se všemi tvary pravidel jsou ekvivalentní.
- Jazyky  $\mathcal{L}_3$  jsou přijímány nedeterministickým konečným automatem (KA).
  - Varianty: nedeterministický KA, deterministický KA a rozšířený KA disponují stejnou vyjadřovací silou.
- Třída konečných jazyků  $\mathcal{L}_{Fin}$  je podmnožinou regulárních jazyků, platí  $\mathcal{L}_{Fin} \subset \mathcal{L}_3$ .
- Příklad jazyků:
  - $\{a^n b^m \mid n \geq 10 \wedge m \leq 20\} \in \mathcal{L}_3 \setminus \mathcal{L}_{Fin}$
  - $\{abc\} \in \mathcal{L}_{Fin}$

## 1.5 Vlastnosti formálních jazyků a jejich rozhodnutelnost

- Které vlastnosti jazyků, resp. tříd jazyků, zkoumáme?
  - **Uzavřenost** vůči nějaké operaci – zda výsledek operace patří do stejné třídy.
  - **Rozhodnutelnost** problému – zda je problém rozhodnutelný, částečně rozhodnutelný a nebo nerozhodnutelný.

### Operace nad jazyky

- Základní množinové operace – sjednocení, průnik, doplněk.
- Základní jazykové operace – konkatenace, iterace, reverzace.
- Substituce
  - Nechť  $\Sigma$  a  $\Gamma$  jsou abecedy. Funce  $\sigma : \Sigma^* \rightarrow 2^{\Gamma^*}$  se nazývá substituce právě tehdy, když
    - \*  $\sigma(\epsilon) = \epsilon$
    - \*  $\forall x, y \in \Sigma^* : \sigma(xy) = \sigma(x)\sigma(y)$
  - Neformálně řečeno, každý symbol vstupního řetězce je nahrazen množinou řetězců.

- Příklad:
  - \* Substituce:  $\sigma(a) = \{0, 00\}$ ,  $\sigma(b) = \{1, 111, 1110\}$
  - \* Aplikace:  $\sigma(ab) = \{0, 00\}\{1, 111, 1110\} = \{01, 0111, 01110, 001, 00111, 001110\}$
- Morfismus
  - Nechť  $\Sigma$  a  $\Gamma$  jsou abecedy. Funce  $\phi : \Sigma^* \rightarrow \Gamma^*$  se nazývá morfismus právě tehdy, když
    - \*  $\phi(\epsilon) = \epsilon$
    - \*  $\forall x, y \in \Sigma^* : \phi(xy) = \phi(x)\phi(y)$
  - Neformálně řečeno, morfismus je zvláštní případ substitutce, kde každý substituovaný jazyk má právě jedno slovo.
  - Příklad:
    - \* Morfismus:  $\phi(a) = 00$ ,  $\phi(b) = 01$ ,  $\phi(c) = 10$
    - \* Aplikace:  $\phi(abac) = \phi(a)$

|                | <b>L3</b> | <b>L2-D</b> | <b>L2</b> | <b>L1</b> | <b>L0-Rec</b> | <b>L0</b> |
|----------------|-----------|-------------|-----------|-----------|---------------|-----------|
| Sjednocení     | Ano       | Ne          | Ano       | Ano       | Ano           | Ano       |
| Pruník         | Ano       | Ne          | Ne        | Ano       | Ano           | Ano       |
| Pruník s L3    | Ano       | Ano         | Ano       | Ano       | Ano           | Ano       |
| Konkatenace    | Ano       | Ne          | Ano       | Ano       | Ano           | Ano       |
| Iterace        | Ano       | Ne          | Ano       | Ano       | Ano           | Ano       |
| Doplnek        | Ano       | Ano         | Ne        | Ano       | Ano           | Ne        |
| Reverzace      | Ano       | Ne          | Ano       | Ano       | Ano           | Ano       |
| Substituce     | Ano       | Ne          | Ano       | Ano       | Ne            | Ano       |
| Morfismus      | Ano       | Ne          | Ano       | Ano       | Ne            | Ano       |
| Inv. morfismus | Ano       | Ano         | Ano       | Ano       | Ano           | Ano       |

Obrázek 1.2: Uzavřenost jednotlivých tříd jazyků vůči operacím.

- Vzhledem k de Morganovým zákonům:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cap \overline{L_2}} = \overline{\overline{L_1} \cup \overline{L_2}}$$

pro operace průnik, sjednocení a doplněk platí, že libovolná jazyková třída je uzavřena bud' na

- doplněk
- a nebo na všechny tři.

- Proč? Neformálně: ze dvou těch operací jsem schopen udělat tu třetí.

## Problémy nad jazyky

- Náležitost / nenáležitost jazyka do třídy jazyků.
- Neprázdnost jazyka – Jazyk je neprázdný, pokud obsahuje aspoň 1 řetězec.
- Prázdnost jazyka – Jazyk je prázdný, pokud neobsahuje žádný řetězec.
- Konečnost jazyka – Jazyk je konečný, pokud obsahuje konečný počet řetězců.
- Náležitost řetězce – Řetězec náleží jazyku, pokud je jeho součástí.
- Univerzalita – Jazyk obsahuje všechny řetězce nad abecedou ( $L = \Sigma^*$ ).
- Inkluze jazyků – Jazyk  $L_1$  je inkluze jazyka  $L_2$ , pokud je jazyk  $L_1$  podmnožinou jazyka  $L_2$ .
- Ekvivalence gramatik – Gramatiky  $G_1$  a  $G_2$  jsou ekvivalentní, pokud platí, že  $L(G_1) = L(G_2)$  ( $G_1$  a  $G_2$  mohou být různé).

|                | <b>L3</b> | <b>L2-D</b> | <b>L2</b> | <b>L1</b> | <b>L0-Rec</b> | <b>L0</b> |
|----------------|-----------|-------------|-----------|-----------|---------------|-----------|
| Neprázdnost    | Ano       | Ano         | Ano       | Castecne  | Castecne      | Castecne  |
| Prazdnost      | Ano       | Ano         | Ano       | Ne        | Ne            | Ne        |
| Konecnost      | Ano       | Ano         | Ano       | Ne        | Ne            | Ne        |
| Naležitost     | Ano       | Ano         | Ano       | Ano       | Ano           | Castecne  |
| Inkluze        | Ano       | Ne          | Ne        | Ne        | Ne            | Ne        |
| Ekvivalence g. | Ano       | Ano         | Ne        | Ne        | Ne            | Ne        |

Obrázek 1.3: Rozhodnutelnost problémů v jednotlivých třídách jazyků.

## 1.6 Vlastnosti regulárních jazyků

Uzávěrové vlastnosti a rozhodnutelnost problémů regulárních jazyků.

### 1.6.1 Sjednocení, průnik, doplněk, iterace, konkatenace

- Uzavřenosť regulárních jazyků vůči operacím sjednocení, průnik, doplněk, iterace a konkatenace vyplývá z definice regulárních výrazů, resp. regulárních množin a ekvivalence regulárních množin a regulárních jazyků (viz otázka regulární výrazy).

- Alternativně např. pro sjednocení (pro ostatní operace je to analogické). Mějme regulární jazyky  $L_1$  a  $L_2$ . Pak musí existovat KA  $M_1$  a  $M_2$ , které je přijímají. Pokud jazyk  $L_3 = L_1 \cup L_2$  je regulární, pak musí existovat KA  $M_3 = M_1 \cup M_2$ , který ho přijímá.

### 1.6.2 Doplněk (*co*)

Necht'  $L \subseteq \Sigma^*$  je regulární jazyk. Necht'  $M = (Q, \Sigma, \delta, q_0, F)$  je úplně definovaný KA, pro který platí  $L = L(M)$ . Pak KA  $M' = (Q, \Sigma, \delta, q_0, Q - F)$  zřejmě přijímá jazyk  $L_{co} = \Sigma^* - L$ , tj. doplněk jazyka  $L$ .

### 1.6.3 Průnik

Uzavřenost vzhledem k průniku plyne z de Morganových zákonů:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cap \overline{L_2}} = \overline{L_1} \cup \overline{L_2}$$

a tedy  $L_1, L_2 \in \mathcal{L}_3 \Leftrightarrow L_1 \cap L_2 \in \mathcal{L}_3$

### 1.6.4 Regulárnost

- Necht'  $L \subseteq \Sigma^*$ , jak řešit problém:  $L \in \mathcal{L}_3$ ?
  - Dokážeme existenci takového KA  $M$ , pro který platí  $L(M) = L$ .
  - S využitím Myhill-Nerodovi věty. Najdeme relaci pravé kongruence s konečným indexem na  $\Sigma^*$  takovou, že jazyk  $L$  je sloučením některých tříd rozkladu.

### 1.6.5 Neregulárnost

- Necht'  $L \subseteq \Sigma^*$ , jak řešit problém:  $L \notin \mathcal{L}_3$ ?
  - S využitím Pumping Lemma pro regulární jazyky.
  - S využitím Myhill-Nerodovi ukážeme, že neexistuje taková relace pravé kongruence s konečným indexem na  $\Sigma^*$  taková, že jazyk  $L$  je sloučením některých tříd rozkladu.

### 1.6.6 Prázdnost/neprázdnost

Problém je rozhodnutelný. Můžu sestrojit algoritmus, který vyzkouší, zda je možné, se nějak dostat z výchozího stavu do nějakého konečného (problém dosažitelnosti v grafu). Formálně: K jazyku  $L \in \mathcal{L}_3$  sestrojíme úplně definovaný DKA  $M$ , takový, že  $L = L(M)$ , pak

$$L(M) \neq \emptyset \Leftrightarrow \exists q \in Q : (q \in F \wedge q \text{ je dosažitelný z } q_0)$$

### 1.6.7 Univerzálnita

Problém je rozhodnutelný. K jazyku  $L \in \mathcal{L}_3$  sestrojíme úplně definovaný DKA  $M$ , takový, že  $L = L(M)$ , pak

$$L(M) = \Sigma^* \Leftrightarrow \forall q \in Q : (q \in F \vee q \text{ je nedosažitelný z } q_0)$$

### 1.6.8 Náležitost

Problém je rozhodnutelný. K jazyku  $L \in \mathcal{L}_3$  sestrojíme úplně definovaný DKA  $M$ , takový, že  $L = L(M)$ , pak

$$w \in L \Leftrightarrow (q_0, w) \vdash^* (q, \epsilon) \wedge q \in F$$

### 1.6.9 Ekvivalence jazyků

Problém je rozhodnutelný. K jazyku  $L_1, L_2 \in \mathcal{L}_3$  sestrojíme úplně definovaný DKA  $M_1, M_2$ , takový, že  $L_1 = L(M_1)$  a  $L_2 = L(M_2)$ . Automaty  $M_1, M_2$  minimalizujeme a porovnáme jejich struktury (ignorujeme identifikátory stavů).

### 1.6.10 Konečnost

- [[todo]]

## 1.7 Vlastnosti bezkontextových jazyků

Uzávěrové vlastnosti a rozhodnutelnost problémů bezkontextových jazyků.

### 1.7.1 Substituce

**Věta 6.2** Třída bezkontextových jazyků je uzavřena vůči substituci.

*Důkaz.*

- Ve shodě s definicí substituce nechť  $\Sigma = \{a_1, a_2, \dots, a_n\}$  je abeceda bezkontextového jazyka  $L$  a  $L_a$  pro  $a \in \Sigma$  libovolné bezkontextové jazyky. Nechť  $G = (N, \Sigma, P, S)$  a  $G_a = (N_a, \Sigma_a, P_a, S_a)$  pro  $a \in \Sigma$  jsou gramatiky, pro které  $L = L(G)$  a  $L_a = L(G_a)$  pro  $a \in \Sigma$ .
- Předpokládejme, že  $N \cap N_a = \emptyset$  a  $N_a \cap N_b = \emptyset$  pro každé  $a, b \in \Sigma, a \neq b$ . Sestrojme gramatiku  $G' = (N', \Sigma', P', S)$  takto:
  1.  $N' = N \cup \bigcup_{a \in \Sigma} N_a$ .
  2.  $\Sigma' = \bigcup_{a \in \Sigma} \Sigma_a$ .
  3. Nechť  $h$  je morfismus na  $N \cup \Sigma$  takový, že
    - $h(A) = A$  pro  $A \in N$  a
    - $h(a) = S_a$  pro  $a \in \Sigma$a nechť  $P' = \{A \rightarrow h(\alpha) \mid (A \rightarrow \alpha) \in P\} \cup \bigcup_{a \in \Sigma} P_a$ .
- Uvažujme libovolnou větu  $a_{i_1}a_{i_2}\dots a_{i_m} \in L$  a věty  $x_j \in L_{a_j}, 1 \leq j \leq m$ . Pak  $S \xrightarrow[G']{*} S_{a_{i_1}}S_{a_{i_2}}\dots S_{a_{i_m}} \xrightarrow[G']{*} x_1S_{a_{i_2}}\dots S_{a_{i_m}} \xrightarrow[G']{*} \dots \xrightarrow[G']{*} x_1x_2\dots x_m$  a tedy  $L' \subseteq L(G')$ .  
Podobně  $L(G') \subseteq L'$ . □

### 1.7.2 Sjednocení, konkatenace, iterace, pozitivní iterace, morfismus

Nechť  $L_a$  a  $L_b$  jsou bezkontextové jazyky.

1. Uzavřenosť vůči  $\cup$  plyne ze substituce  $L_a, L_b$  do jazyka  $\{a, b\}$ .
2. Uzavřenosť vůči  $.$  plyne ze substituce  $L_a, L_b$  do jazyka  $\{ab\}$ .
3. Uzavřenosť vůči  $*$  plyne ze substituce  $L_a$  do jazyka  $\{a\}^*$ .
4. Uzavřenosť vůči  $+$  plyne ze substituce  $L_a$  do jazyka  $\{a\}^+$ .
5. Nechť  $h$  je daný morfismus a  $L'_a = \{h(a)\}$  pro  $a \in \Sigma$ . Substitucí jazyků  $L'_a$  do jazyka  $L$  získáme jazyk  $h(L)$ .

### 1.7.3 Průnik

- Bezkontextové jazyky nejsou uzavřeny vůči průniku, protipříklad:

$$L_1 = \{a^n, b^n, c^m \mid n, m \geq 1\} \in \mathcal{L}_2$$

$$L_2 = \{a^m, b^n, c^n \mid n, m \geq 1\} \in \mathcal{L}_2$$

$$L_1 \cap L_2 = \{a^n, b^n, c^n \mid n \geq 1\} \notin \mathcal{L}_2$$

### 1.7.4 Doplňek

- Bezkontextové jazyky nejsou uzavřeny vůči doplňku.
- Předpokládejme, že bezkontextové jazyky jsou uzavřeny vůči doplňku. Potom z uzavřenosťi vůči sjednocení a de Morganových zákonů:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cap \overline{L_2}} = \overline{\overline{L_1} \cup \overline{L_2}}$$

vyplývá uzavřenosť vůči průniku. Což je spor.

### 1.7.5 Bezkontextovost

- Nechť  $L \subseteq \Sigma^*$ , jak řešit problém:  $L \in \mathcal{L}_2$ ?
  - Dokážeme existenci takového NZA  $M$ , pro který platí  $L(M) = L$ .

### 1.7.6 Nebezkontextovost

- Nechť  $L \subseteq \Sigma^*$ , jak řešit problém:  $L \notin \mathcal{L}_3$ ?
  - S využitím Pumping Lemma pro bezkontextové jazyky.

### 1.7.7 Prázdnost/neprázdnost

Zásobníkový automat je příliš komplexní model pro řešení takovýchto problémů, radši budeme vycházet z bezkontextových gramatik. K rozhodování neprázdnosti lze využít algoritmus iterativně určující množinu  $N_t$  nonterminálů generujících terminální řetězce (viz přednášky). Pak  $L(G) \neq \emptyset \Leftrightarrow S \in N_t$ .

### 1.7.8 Náležitost

U problému příslušnosti řetězce můžeme např. určit průnik NZA s KA přijímacím právě řetězec  $w$  a pak ověřit neprázdnost (převod na gramatiku, problém prázdnosti gramatiky umíme řešit).

### 1.7.9 Konečnost

Problém konečnosti můžeme rozhodovat na základě platnosti Pumping lemma pro CFL:

- Dle Pumping lemma pro bezkontextové jazyky existuje pro každý bezkontextový jazyk  $L$  konstanta  $k \in \mathbb{N}$  taková, že každou větu  $w \in L$ ,  $|w| \geq k$ , můžeme rozepsat jako  $uvwxy$ , kde  $vx \neq \varepsilon$  a  $|vwx| \leq k$ , a  $\forall i \in \mathbb{N} : uv^iwx^i y \in L$ .
- Pro testování konečnosti tedy postačí ověřit, že žádný řetězec ze  $\Sigma^*$  o délce mezi  $k$  a  $2k - 1$  nepatří do daného jazyka:
  - Pokud takový řetězec existuje, může být „napumpován“ a dostaváme nekonečně mnoho řetězců patřících do daného jazyka.
  - Jestliže takový řetězec neexistuje,  $k - 1$  je horní limit délky řetězců  $L$ .
  - Pokud by existoval řetězec délky  $2k$  nebo větší patřící do  $L$ , můžeme v něm podle Pumping lemma najít  $vwx$  a vypustit  $vx$ . Vzhledem k tomu, že  $0 < |vx| \leq k$ , postupným opakováním vypouštění bychom se dostali k nutné existenci řetězce z  $L$  o délce mezi  $k$  a  $2k - 1$ .
- K určení konstanty  $k$  postačí reprezentovat  $L$  pomocí bezkontextové gramatiky v CNF s  $n$  nonterminály a zvolit  $k = 2^n$  (viz důkaz Pumping lemma).

□

### 1.7.10 Ekvivalence jazyků

- Nerozhodnutelný problém, dokonce i sama inkluze je nerozhodnutelná.
- Důkaz redukcí, komplikovanější.

## 1.8 Vlastnosti kontextových jazyků

Uzávěrové vlastnosti a rozhodnutelnost problémů kontextových jazyků.

### 1.8.1 Sjednocení, průnik, konkatenace, iterace, komplement

**Věta 8.10** Třída kontextových jazyků je uzavřena vůči operacím  $\cup$ ,  $\cap$ ,  $.$ ,  $*$  a komplementu.

*Důkaz.*

- Uzavřenosť vůči  $\cup$ ,  $\cap$ ,  $.$  a  $*$  lze ukázat stejně jako u rekurzívnych spočetných jazyků.
- Důkaz uzavřenosťi vůči komplementu je značne komplikovaný (všimněme si, že LOA je nedeterministický a nelze tudíž užít konstrukce použité u rekurzívnych jazyků) – zájemci naleznou důkaz v doporučené literatuře.

□

❖ Poznamenejme, že již víme, že u kontextových jazyků

- lze rozhodovat členství věty do jazyka (rekurzivnost) a
- nelze rozhodovat inkluzi jazyků (neplatí ani pro bezkontextové jazyky).

❖ Dále lze ukázat, že pro kontextové jazyky nelze rozhodovat prázdnost jazyka (užije se redukce z Postova problému přiřazení – viz další přednášky).

## 1.9 Vlastnosti obecných jazyků

Uzávěrové vlastnosti a rozhodnutelnost problémů obecných jazyků.

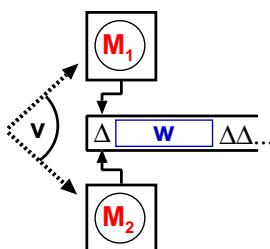
- Zkoumáme pro rekurzivně vyčíslitelné  $\mathcal{L}_{RE}$  a rekurzivní  $\mathcal{L}_{Rec}$  zároveň.

### 1.9.1 Sjednocení, průnik, konkatenace, iterace

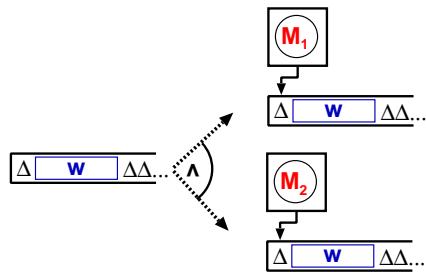
**Věta 8.4** Třídy rekurzívnych a rekurzivně vyčíslitelných jazyků jsou uzavřeny vůči operacím  $\cup$ ,  $\cap$ ,  $.$  a  $*$ .

*Důkaz.* Nechť  $L_1, L_2$  jsou jazyky přijímané TS  $M_1, M_2$ . Zřejmě můžeme předpokládat, že množiny stavů TS  $M_1, M_2$  jsou disjunktní.

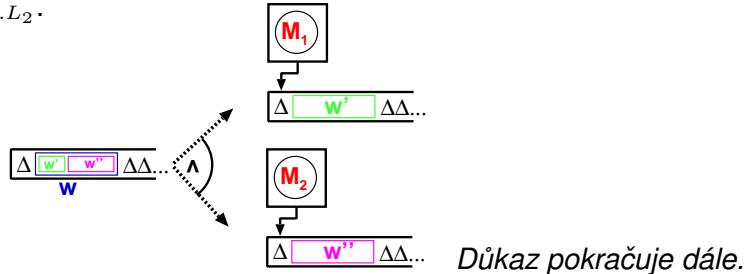
- NTS  $M_{L_1 \cup L_2}, L(M_{L_1 \cup L_2}) = L_1 \cup L_2$ , sestrojíme tak, že sjednotíme po složkách stroje  $M_1$  a  $M_2$ , zavedeme nový počáteční stav, z něj nedeterministické přechody přes  $\Delta/\Delta$  do obou původních počátečních stavů a sloučíme původní koncové stavы do jediného nového koncového stavu.



- Třípáskový TS  $M_{L_1 \cap L_2}$ ,  $L(M_{L_1 \cap L_2}) = L_1 \cap L_2$ , okopíruje vstup z první pásky na druhou, na ní simuluje stroj  $M_1$ , pokud ten přijme, okopíruje vstup z první pásky na třetí, na ní simuluje stroj  $M_2$ , pokud i ten přijme, přijme i stroj  $M_{L_1 \cap L_2}$ .



- Třípáskový NTS  $M_{L_1 \cdot L_2}$ ,  $L(M_{L_1 \cdot L_2}) = L_1 \cdot L_2$ , okopíruje nedeterministicky zvolený prefix vstupu z první pásky na druhou, na ní simuluje stroj  $M_1$ , pokud ten přijme, okopíruje zbytek vstupu z první pásky na třetí, na ní simuluje stroj  $M_2$ , pokud i ten přijme, přijme i stroj  $M_{L_1 \cdot L_2}$ .



- Dvoupáskový NTS  $M_{L_1^*}$ ,  $L(M_{L_1^*}) = L_1^*$ , je zobecněním předchozího stroje: po částečném kopírování vstupu z první pásky na druhou a na ní simuluje opakování stroje  $M_1$ . Obsah druhé pásky má ohraničený speciálními značkami a po každé simulaci stroje  $M_1$  ho smaže. Umožňuje samozřejmě posuv pravé značky dále doprava při nedostatku místa.

Jsou-li stroje  $M_1$  a  $M_2$  úplné, je možné vybudovat stroje podle výše uvedených pravidel také jako **úplné** (u  $M_{L_1 \cup L_2}$ ,  $M_{L_1 \cap L_2}$ ,  $M_{L_1 \cdot L_2}$  je to okamžité, u  $M_{L_1^*}$  nepřipustíme načítání prázdného podřetězce vstupu z 1. na 2. pásku – pouze umožníme jednorázově přjmout prázdný vstup). To dokazuje uzavřenosť vůči uvedeným operacím také u **rekurzívnych jazyků**.

□

### 1.9.2 Doplněk

**Věta 8.5** Třída rekurzívních jazyků je uzavřena vůči komplementu.

*Důkaz.* TS  $M$  přijímající rekurzívní jazyk  $L$  vždy zastaví. Snadno upravíme  $M$  na  $M'$ , který při nepřijetí řetězce vždy přejde do unikátního stavu  $q_{reject}$ . TS  $\overline{M}$ ,  $L(\overline{M}) = \overline{L}$ , snadno dostaneme z  $M'$  záměnou  $q_F$  a  $q_{reject}$ .  $\square$

❖ Třída rekurzívne vyčíslitelných jazyků není uzavřena vůči komplementu!

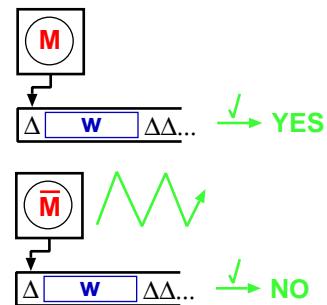
- Výše uvedené konstrukce nelze užít – cyklení zůstane zachováno.
- Důkaz neuzavřenosti bude uveden v dalších přednáškách.

**Věta 8.6** Jsou-li  $L$  i  $\overline{L}$  rekurzívne vyčíslitelné, pak jsou oba rekurzívní.

*Důkaz.*

Mějme  $M$ ,  $L(M) = L$ , a  $\overline{M}$ ,  $L(\overline{M}) = \overline{L}$ . Úplný TS přijímající  $L$  sestrojíme takto:

- Použijeme dvě pásky. Na jedné budeme simulovat  $M$ , na druhé  $\overline{M}$ . Simulace se bude provádět proloženě krok po kroku: krok  $M$ , krok  $\overline{M}$ , krok  $M$ , ...
- Přijmeme, právě když by přijal  $M$ , zamítneme abnormálním zastavením, právě když by přijal  $\overline{M}$ . Jedna z těchto situací určitě nastane v konečném počtu kroků.



Existence úplného TS pro  $\overline{L}$  plyne z uzavřenosti rekurzívních jazyků vůči komplementu.  $\square$

❖ Důsledkem výše uvedených vět je mj. to, že pro  $L$  a  $\overline{L}$  musí vždy nastat jedna z následujících situací:

- $L$  i  $\overline{L}$  jsou rekurzívní,
- $L$  ani  $\overline{L}$  nejsou rekurzívne vyčíslitelné,
- jeden z těchto jazyků je rekurzívne vyčíslitelný, ale ne rekurzívní, druhý není rekurzívne vyčíslitelný.

## Kapitola 2

**TIN – Konečné automaty (jazyky přijímané KA, varianty KA, minimalizace KA, Mihill-Nerodova věta).**

### 2.1 Zdroje

- tin\_2021\_merged.pdf
- TIN\_2020-09-22.mp4
- TIN\_2020-09-29.mp4
- TIN\_2020-10-06.mp4

### 2.2 Konečný automat

Konečné automaty dokáží přijímat regulární jazyky.

**Definice** Konečný automat (KA) je pětice  $M = (Q, \Sigma, \delta, q_0, F)$ , kde

- $Q$  je konečná množina stavů;
- $\Sigma$  je vstupní abeceda;
- $\delta$  je přechodová funkce (parciální funkce),
  - $\delta : Q \times \Sigma \rightarrow 2^Q$ ;
- $q_0$  je výchozí stav,
  - $q_0 \in Q$ ;
- $F$  je množina koncových stavů,
  - $F \subseteq Q$ ;

**Konfigurace** Konfigurace KA je dvojice  $(q, w) \in Q \times \Sigma^*$ , kde

- $q$  je aktuální stav;
- $w$  je nezpracovaná část vstupního řetězce.

**Počáteční konfigurace** Počáteční konfigurace je taková konfigurace  $(q, w)$ , kde  $w \in \Sigma^*$  je vstupní řetězec a  $q \in Q$  je výchozí stav.

**Finální konfigurace** Finální konfigurace je taková konfigurace  $(q, w)$ , kde  $w = \epsilon$  je  $q \in F$ .

**Přechod** Přechod (krok výpočtu) KA je binární relace (značíme  $\vdash$ ) na množině konfigurací  $\vdash \subseteq (Q \times \Sigma^*)^2$ , taková, že

$$(q, w) \vdash (q', w') \Leftrightarrow \exists a \in \Sigma : w = aw' \wedge q' \in \delta(q, a)$$

**Jazyk přijímaný** Mějme KA  $M = (Q, \Sigma, \delta, q_0, F)$  a jazyk  $L(M)$ , který je přijímaný KA  $M$ .

$$L(M) = \{w \mid w \in \Sigma^* \wedge (q_0, w) \vdash^* (q_f, \epsilon) \wedge q_f \in F\}$$

**Dosažitelné a nedosažitelné stavy** Mějme KA  $M = (Q, \Sigma, \delta, q_0, F)$ . Stav  $q \in Q$  je dosažitelný, pokud platí  $\exists w \in \Sigma^* : (q_0, w) \vdash^* (q, \epsilon)$ . Stav je nedosažitelný, pokud není dosažitelný.

**Relace nerozlišitelnosti** Relace nerozlišitelnosti je binární relace nad množinou stavů. Neformálně říká, že z jednoho ze stavů, automat dokáže akceptovat rozlišovací řetězec, z druhého nikoliv. Formálně: nechť  $p, q \in Q$  jsou rozlišitelné, pak platí

$$\forall w \in \Sigma^* : (p, w) \vdash^* (p', \epsilon) \wedge (q, w) \vdash^* (q', \epsilon) \wedge ((p' \in F \wedge q' \notin F) \vee (p' \notin F \wedge q' \in F))$$

Dva stavy jsou nerozlišitelné, pokud nejsou rozlišitelné.

**k-nerozlišitelnost** Říkáme, že stavy  $q_1, q_2 \in Q$  jsou k-nerozlišitelné a píšeme  $q_1 \equiv^k q_2$ , právě když neexistuje  $w \in \Sigma^*, |w| \leq k$ , který rozlišuje  $q_1$  a  $q_2$ . Lze dokázat, že  $\equiv$  je relací akvivalence na  $Q$  (reflexivní, symetrická, tranzitivní). Neformálně relace říká, jak dlouhé slovo dané stavy rozlišuje.

## 2.3 Varianty konečného automatu

Všechny varianty konečného automatu mají stejnou vyjadřovací sílu (jsou mezi sebou převoditelné).

**Nedeterministický konečný automat** Nedeterministický konečný automat (NKA) je výchozí konečný automat ( $NKA = KA$ ).

**Deterministický konečný automat** Deterministický konečný automat (DKA) se od NKA liší pouze tvarem přechodové funkce. Formálně:

- $\delta$  je přechodová funkce (parciální funkce),
  - $\delta : Q \times \Sigma \rightarrow Q$

**Rozšířený konečný automat** Rozšířený (nedeterministický) konečný automat (RKA) se od NKA liší pouze tvarem přechodové funkce. Rozšiřuje ji o tzv. epsilon přechody. Formálně:

- $\delta$  je přechodová funkce (parciální funkce),
  - $\delta : Q \times (\Sigma \cup \{\epsilon\}) \rightarrow 2^Q$

**Úplně definovaný konečný automat** Úplně definovaný konečný automat se od NKA liší pouze tvarem přechodové funkce. Formálně:

- $\delta$  je přechodová funkce (totální funkce),

$$-\ \delta : Q \times \Sigma \rightarrow 2^Q$$

$$-\ \forall q \in Q \ \forall a \in \Sigma : \delta(q, a) \neq \emptyset$$

**Redukovaný úplně definovaný deterministický konečný automat** Úplně definovaný DKA nazýváme redukovaný (také minimální), jestliže žádný  $q \in Q$  není nedosažitelný a žádná dvojice  $p, q \in Q$  není nerozlišitelná.

## 2.4 Determinizace NKA

### Algoritmus

Vstup: NKA  $M = (Q, \Sigma, \delta, q_0, F)$

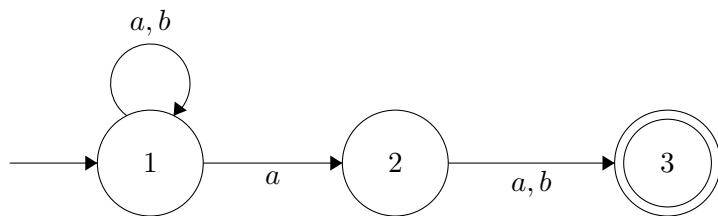
Výstup: DKA  $M' = (Q', \Sigma, \delta', q'_0, F')$ ,  $L(M) = L(M')$

Metoda:

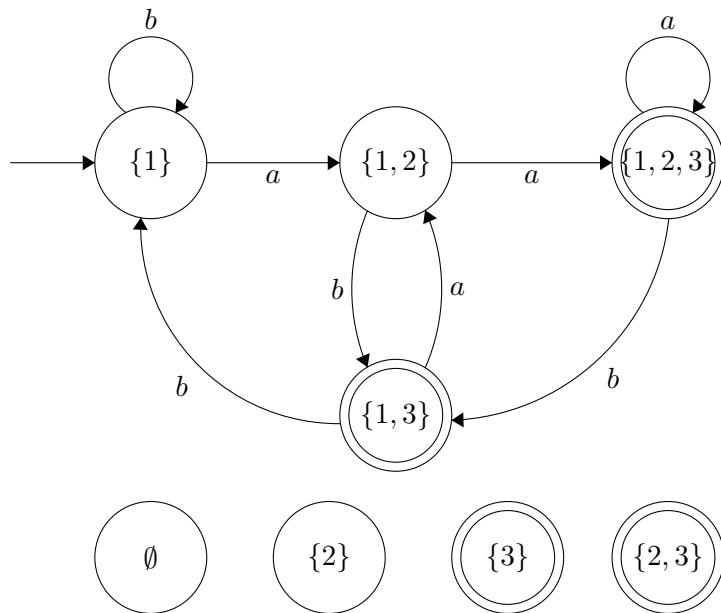
1. Polož  $Q' = 2^Q$ .
2. Polož  $q'_0 = \{q_0\}$ .
3. Polož  $F' = \{S \mid S \in 2^Q \wedge S \cap F \neq \emptyset\}$ .
4. Pro všechna  $S \in 2^Q$  a pro všechna  $a \in \Sigma$  polož:
  - $\delta'(S, a) = \bigcup_{q \in S} \delta(q, a)$ .

Obrázek 2.1: Algoritmus determinizace NKA.

### Příklad



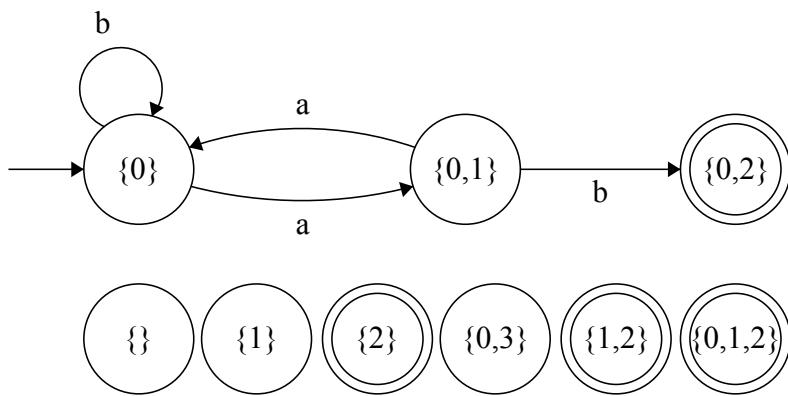
Obrázek 2.2: Nedeterministický konečný automat  $M$ .



Obrázek 2.3: Deterministický konečný automat  $M'$ ,  $L(M) = L(M')$ .

## 2.5 Minimalizace konečného automatu

- Proč? Zrychlení vykonávání konečného automatu.
- Jak? Odstraněním nepotřebných přechodů a zmenšením počtu stavů.
- Postup (předpokládá DKA):
  1. Odstranění nedosažitelných stavů.
  2. Převod DKA na úplně definovaný DKA.
  3. Převod úplně definovaný DKA na redukovaný DKA (odstranění nerozlišitelných stavů).
  4. Odstranění *sink* stavu.



Obrázek 2.4: Deterministický konečný automat.

### 2.5.1 Eliminace nedosažitelných stavů

**Algoritmus 3.4** Eliminace nedosažitelných stavů

**Vstup:** Deterministický konečný automat  $M = (Q, \Sigma, \delta, q_0, F)$ .

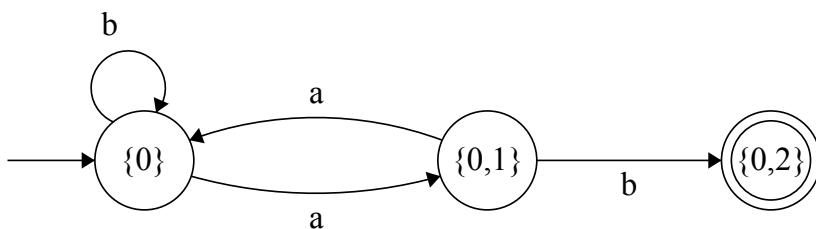
**Výstup:** Deterministický konečný automat  $M'$  bez nedosažitelných stavů,  $L(M) = L(M')$ .

**Metoda:**

1.  $i := 0$
2.  $S_i := \{q_0\}$
3. **repeat**
4.    $S_{i+1} := S_i \cup \{q \mid \exists p \in S_i \ \exists a \in \Sigma : \delta(p, a) = q\}$
5.    $i := i + 1$
6. **until**  $S_i = S_{i-1}$
7.  $M' := (S_i, \Sigma, \delta|_{S_i}, q_0, F \cap S_i)$

Základem algoritmu minimalizace deterministického konečného automatu je koncept nerozlišitelných stavů.

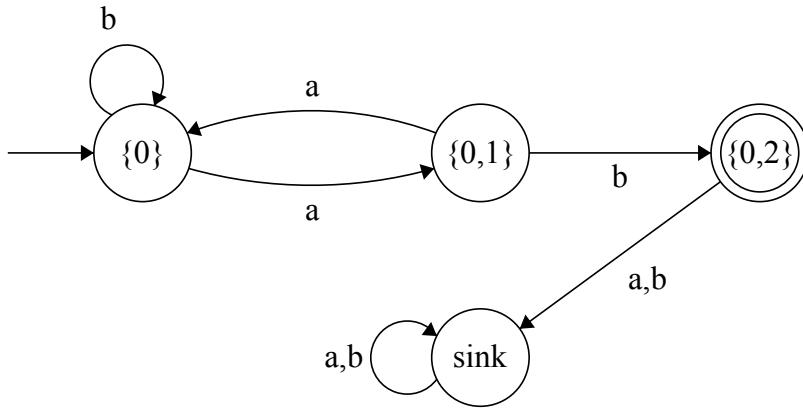
Obrázek 2.5: Algoritmus eliminace nedosažitelných stavů.



Obrázek 2.6: Deterministický konečný automat bez nedosažitelných stavů.

### 2.5.2 Převod DKA na úplně definovaný

- Vstup: DKA  $M = (Q, \Sigma, \delta, q_0, F)$
- Výstup: úplně definovaný DKA  $M'$
- Metoda:
  1.  $\forall q \in Q \ \forall a \in \Sigma : \delta(q, a) = \emptyset \Rightarrow \delta'(q, a) = q_s$
  2.  $\forall a \in \Sigma : \delta'(q_s, a) = q_s$
  3.  $M' = (Q \cup \{q_s\}, \Sigma, \delta \cup \delta', q_0, F)$
- $q_s$  je tzv. *sink* stav.



Obrázek 2.7: Úplně definovaný deterministický konečný automat.

### 2.5.3 Převod úplně definovaného DKA na redukovaný

**Algoritmus 3.5** Převod na redukovaný deterministický konečný automat.

**Vstup:** Úplně definovaný DKA  $M = (Q, \Sigma, \delta, q_0, F)$ .

**Výstup:** Redukovaný DKA  $M' = (Q', \Sigma, \delta', q'_0, F')$ ,  $L(M) = L(M')$ .

**Metoda:**

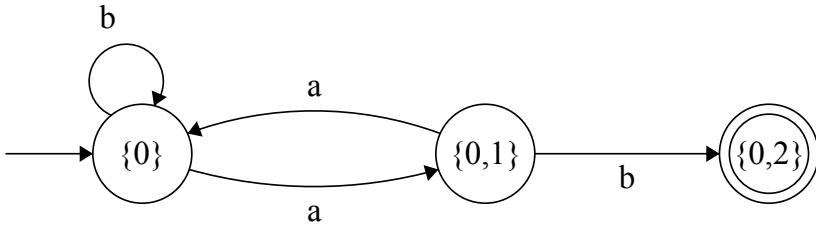
1. Odstraň nedosažitelné stavy s využitím algoritmu 3.4.
2.  $i := 0$
3.  $\stackrel{0}{\equiv} := \{(p, q) \mid p \in F \iff q \in F\}$
4. **repeat**
5.  $\stackrel{i+1}{\equiv} := \{(p, q) \mid p \stackrel{i}{\equiv} q \wedge \forall a \in \Sigma : \delta(p, a) \stackrel{i}{\equiv} \delta(q, a)\}$
6.  $i := i + 1$
7. **until**  $\stackrel{i}{\equiv} = \stackrel{i-1}{\equiv}$
8.  $Q' := Q / \stackrel{i}{\equiv}$
9.  $\forall p, q \in Q \ \forall a \in \Sigma : \delta'([p], a) = [q] \Leftrightarrow \delta(p, a) = q$
10.  $q'_0 = [q_0]$
11.  $F' = \{[q] \mid q \in F\}$

**Poznámka 3.2** Výraz  $[x]$  značí ekvivalentní třídu určenou prvkem  $x$ .

Obrázek 2.8: Algoritmus převodu úplně definovaného DKA na redukovaný.

### 2.5.4 Odstranění *sink* stavu

- Vstup: DKA  $M = (Q, \Sigma, \delta, q_0, F)$
- Výstup: DKA  $M'$  bez *sink* stavu
- Metoda:
  1.  $Q' = \{q \mid q \in Q \wedge \exists w \in \Sigma^* \wedge q_f \in F \wedge (q, w) \vdash^* (q_f, \epsilon)\}$
  2.  $\delta' = \{(p, a, q) \mid \delta(p, a) = q \wedge q \notin Q - Q'\}$
  3.  $M' = (Q', \Sigma, \delta', q_0, F)$



Obrázek 2.9: Úplně definovaný deterministický konečný automat bez *sink* stavu.

## 2.6 Myhill-Nerodova věta

**Pumping Lemma** Pumping Lemma poskytuje nutnou podmínu, ale nikoliv dostačující pro regulární jazyky, tj. když jazyk je regulární, tak podmína musí platit. To znamená, že pomocí PL můžeme dokázat, že jazyk není regulární. Ale nemůžeme použít PL k důkazu regularity jazyky. Myšlenka Pumping Lemma spočívá v tom, že pokud mám  $w \in L$ , jehož délka je větší, než je počet stavů KA, který daný jazyk přijímá, tak tam musí být někde smyčka. A tím pádem, můžu tu smyčku bud' úplně vypustit a nebo zopakovat vícekrát.

**Mihill-Nerodova věta – kontext** Mihill-Nerodova věta charakterizuje jazyky nad  $\Sigma^*$  a jejich vztah ke konečným automatům. Je silnější než Pumping Lemma, protože poskytuje nutné a dosačující podmínky pro to, aby jazyk byl regulární. Pomocí ní můžeme dokázat, že jazyk je a nebo není regulární.

**Relace ekvivalence** Relace ekvivalence je binární relace (značíme  $\sim$ ), která je reflexivní, symetrická a tranzitivní. Rozkládá množinu nad kterou je definovaná na tzv. **třídy ekvivalence**. Pro každé dvojice tříd ekvivalence platí, že jsou vzájemně disjunktní. Sloučením všech tříd ekvivalence dostaneme původní množinu. Počet tříd rozkladu je tzv. **index ekvivalence**. Tříd může být i nekonečno, pak index definujeme jako  $\infty$ . Relaci ekvivalence nad množinou  $\Sigma^*$  zapisujeme jako  $\Sigma^* / \sim$ .

**Pravá kongruence** Pravá kongruence (pravá invariance) je speciální typ relace ekvivalence, která splňuje požadovanou vlastnost. Necht'  $\Sigma$  je abeceda a  $\sim$  je relace ekvivalence nad  $\Sigma^*$ . Relace ekvivalence  $\sim$  je pravou kongruencí pokud platí:

$$\forall u, v, w \in \Sigma^* : u \sim v \Rightarrow uw \sim vw$$

Myšlenka: Pro všechny slova  $u, v$  z jazyka platí, že pokud jsem se přes ně dostal do nějakého (stejného) stavu, tak když k oboum připojím slovo  $w$ , tak se zase dostanu do stejného stavu.

**Prefixová ekvivalence** Necht'  $L$  je libovolný (ne nutně regulární) jazyk nad abecedou  $\Sigma$ . Na množině  $\Sigma^*$  definujeme relaci  $\sim_L$  zvanou prefixová ekvivalence pro  $L$  takto:

$$u \sim_L v \Leftrightarrow (\forall w \in \Sigma^* : uw \in L \Leftrightarrow uw \notin L)$$

Myšlenka: slova  $u$  a  $v$  dovedou konečný automat do stejného stavu.

**Mihill-Nerodova věta** Jazyk  $L \subseteq \Sigma^*$  je regulární právě když:

- $\exists \text{KA } M : L(M) = L$
- Existuje pravá kongruence  $\sim$  konečného indexu na  $\Sigma^*$  taková, že  $L$  je sjednocením vybraných tříd rozkladu  $\Sigma^*/\sim$ .
- Relace prefixové ekvivalence  $\sim_L$  má konečný index (existuje minimální KA, který odpovídá).

Pokud dokážu, že nad daným jazykem neexistuje relace pravé kongruence s konečným indexem, tak jazyk není regulární.

# Kapitola 3

## TIN – Regulární množiny, regulární výrazy a rovnice nad regulárními výrazy.

### 3.1 Zdroje

- [tin\\_2021\\_merged.pdf](#)
- [TIN\\_2020-09-29.mp4](#)

### 3.2 Regulární množiny

Regulární jazyky lze definovat pomocí regulárních gramatik, konečných automatů a regulárních množin.

**Definice** Necht'  $\Sigma$  je konečná abeceda. Regulární množiny nad  $\Sigma$  definujeme rekurzivně takto:

- $\emptyset$  je regulární množina nad  $\Sigma$ ;
- $\{\epsilon\}$  je regulární množina nad  $\Sigma$ ;
- $\{a\}$  je regulární množina nad  $\Sigma$  pro  $\forall a \in \Sigma$ ;
- jsou-li  $P$  a  $Q$  regulární množiny nad  $\Sigma$ , pak také
  - $P \cup Q$ ,
  - $P . Q$ ,
  - $P^*$

jsou regulární množiny nad  $\Sigma$ .

- Žádné jiné množiny, než ty, které lze získat pomocí výše uvedených pravidel, nejsou regulárními množinami.

**Příklad (a)** Jazyk definovaný regulární množinou.

$$L_{RM} = (\{a\} \cup \{d\}) . (\{b\}^*) . \{c\}$$

### 3.3 Regulární výrazy

Regulární výrazy jsou pouze zkrácený zápis regulárních množin.

**Definice** Regulární výrazy nad  $\Sigma$  a regulární množiny, které označují, jsou rekurzivně definovány takto:

- $\emptyset$  je regulární výraz označující regulární množinu  $\emptyset$ ;
- $\epsilon$  je regulární výraz označující regulární množinu  $\{\epsilon\}$ ;
- $a$  je regulární výraz označující regulární množinu  $\{a\}$  pro  $\forall a \in \Sigma$ ;
- jsou-li  $p$  a  $q$  regulární výrazy označující regulární množiny  $P$  a  $Q$ , pak
  - $(p + q)$  je regulární výraz označující regulární množinu  $P \cup Q$ ,
  - $(pq)$  je regulární výraz označující regulární množinu  $P \cdot Q$ ,
  - $(p^*)$  je regulární výraz označující regulární množinu  $P^*$ .

**Příklad (b)** Jazyk definovaný regulárním výrazem, platí  $L_{RM} = L_{RV}$ .

$$L_{RV} = (a + d)b^*c$$

**Příklad (c)** Jazyk definovaný regulární gramatikou  $G$ , platí  $L_{RM} = L_{RV} = L(G)$ .

$$G = (\{S, A\}, \{a, b, c, d\}, P, S)$$

$$P = \{S \rightarrow aA \mid dA, A \rightarrow bA \mid c\}$$

**Příklad (d)** Jazyk definovaný konečným automatem  $M$ , platí  $L_{RM} = L_{RV} = L(G) = L(M)$ .

$$M = (\{q_0, q_1, q_2\}, \{a, b, c, d\}, \delta, q_0, \{q_2\})$$

$$\delta(q_0, a) = q_1, \delta(q_0, d) = q_1, \delta(q_1, b) = q_1, \delta(q_1, c) = q_2$$

### 3.4 Rovnice nad regulárními výrazy

**Definice** Rovnice, jejíž složkami jsou koeficienty a neznámé, které reprezentují (dané a hledané) regulární výrazy, nazýváme rovnicemi nad regulárními výrazy.

**Příklad** Uvažujme rovnici nad regulárními výrazy nad abecedou  $\Sigma = \{a, b\}$

$$X = aX + b$$

- Řešením je regulární výraz  $X = a^*b$
- Důkaz:
  - $LS = a^*b$
  - $PS = a(a^*b) + b = a^+b + b = (a^+ + \epsilon)b = a^*b$

## **Shrnutí**

- Ne vždy existuje jediné řešení rovnice nad regulárními výrazy.
- Regulární gramatiky, nedeterministické/deterministické/rozšířené konečné automaty a regulární výrazy mají ekvivalentní vyjadřovací sílu.

## Kapitola 4

# TIN – Zásobníkové automaty (jazyky přijímané ZA, varianty ZA).

### 4.1 Zdroje

- tin\_2021\_merged.pdf
- TIN\_2020-10-06.mp4
- TIN\_2020-10-13.mp4

### 4.2 Zásobníkový automat

Zásobníkové automaty dokáží přijímat bezkontextové jazyky.

**Definice** Zásobníkový automat (ZA) je sedmice  $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ , kde

- $Q$  je konečná množina stavů;
- $\Sigma$  je vstupní abeceda;
- $\Gamma$  je zásobníková abeceda;
- $\delta$  je přechodová funkce (parciální funkce),
  - $\delta : Q \times (\Sigma \cup \{\epsilon\}) \times \Gamma \rightarrow 2^{Q \times \Gamma^*}$ ;
- $q_0$  je výchozí stav,
  - $q_0 \in Q$ ;
- $Z_0$  je výchozí symbol na zásobníku,
  - $Z_0 \in \Gamma$ ;
- $F$  je množina koncových stavů,
  - $F \subseteq Q$ ;

**Konfigurace** Konfigurace ZA je trojice  $(q, w, \alpha) \in Q \times \Sigma^* \times \Gamma^*$ , kde

- $q$  je aktuální stav;
- $w$  je nezpracovaná část vstupního řetězce;
- $\alpha$  je obsah zásobníku.

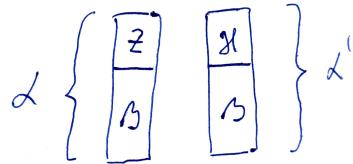
**Počáteční konfigurace** Počáteční konfigurace ZA je taková konfigurace  $(q, w, \alpha)$ , kde  $w \in \Sigma^*$  je vstupní řetězec,  $q \in Q$  je výchozí stav a  $\alpha = Z \in \Gamma$  je výchozí symbol na zásobníku.

**Koncová konfigurace** Koncová konfigurace ZA je taková konfigurace  $(q, w, \alpha)$ , kde  $q \in F$ ,  $w = \epsilon$  a  $\alpha \in \Gamma^*$ .

**Přechod** Přechod (krok výpočtu) ZA je binární relace (značíme  $\vdash$ ) na množině konfigurací  $\vdash \subseteq (Q \times \Sigma^* \times \Gamma^*)^2$ , taková, že

$$(q, w, \alpha) \vdash (q', w', \alpha') \Leftrightarrow \exists a \in (\Sigma \cup \epsilon), \exists Z \in \Gamma, \exists \beta, \gamma \in \Gamma^* :$$

$$: w = aw' \wedge \alpha = Z\beta \wedge \alpha' = \gamma\beta \wedge (q', \gamma) \in \delta(q, a, Z)$$



Obrázek 4.1: Stav zásobníku během přechodu ZA. Platí  $Z \in \Gamma$ ,  $\alpha, \alpha', \beta, \gamma \in \Gamma^*$ .

**Jazyk přijímaný** Mějme ZA  $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$  a jazyk  $L(M)$ , který je přijímaný ZA  $M$ .

$$L(M) = \{w \mid w \in \Sigma^* \wedge (q_0, w, Z_0) \vdash^* (q_f, \epsilon, \epsilon) \wedge q_f \in F\}$$

Avšak existují různé varianty. ZA může přijímat i v  $q_f \in F$ : pouze při vyprázdnění vstupu, nebo pouze při vyprázdnění zásobníku a nebo při vyprázdnění vstupu a vyprázdnění zásobníku.

### 4.3 Varianty zásobníkového automatu

Zde je změna oproti konečným automatům. Všechny varianty zásobníkového automatu nemají stejnou vyjadřovací sílu (nejsou mezi sebou převoditelné).

#### 4.3.1 Nedeterministický zásobníkový automat

Nedeterministický zásobníkový automat (NZA) je výchozí zásobníkový automat (NZA = ZA).

#### 4.3.2 Rozšířený zásobníkový automat

Rozšířený (nedeterministický) zásobníkový automat (RNZA) disponuje stejnou vyjadřovací silou jako NZA (jsou mezi sebou převoditelné).

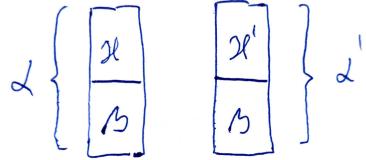
**Definice** RNZA se od NZA liší tvarem přechodové funkce, umožňuje ze zásobníku konzumovat více než 1 symbol. Formálně:

- $\delta$  je přechodová funkce (parciální funkce),
  - $\delta : Q \times (\Sigma \cup \{\epsilon\} \times \Gamma^*) \rightarrow 2^{Q \times \Gamma^*}$

**Konfigurace, přechod** Konfigurace RNZA a NZA jsou shodné. Přechod se liší díky jinému tvaru přechodové funkce. Formálně: Přechod RNZA je binární relace (značíme  $\vdash$ ) na množině konfigurací  $\vdash \subseteq (Q \times \Sigma^* \times \Gamma^*)^2$ , taková, že

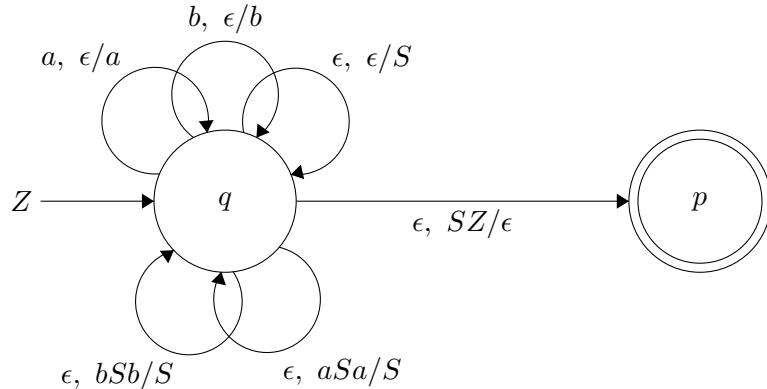
$$(q, w, \alpha) \vdash (q', w', \alpha') \Leftrightarrow \exists a \in (\Sigma \cup \epsilon), \exists \beta, \gamma, \gamma' \in \Gamma^* :$$

$$: w = aw' \wedge \alpha = \gamma\beta \wedge \alpha' = \gamma'\beta \wedge (q', \gamma') \in \delta(q, a, \gamma)$$



Obrázek 4.2: Stav zásobníku během přechodu RNZA. Platí  $\alpha, \alpha', \beta, \gamma, \gamma' \in \Gamma^*$ .

**Příklad** Příklad RNZA, který přijímá jazyk  $L = \{ww^R \mid w \in \{a, b\}^+\}$ . Platí  $L \in \mathcal{L}_2 \setminus \mathcal{L}_{DCF}$ .



Obrázek 4.3: Příklad rozšířeného zásobníkového automatu.

### 4.3.3 Deterministický zásobníkový automat

Deterministický zásobníkový automat (DZA) disponuje menší vyjadřovací sílou než NZA (resp. RNZA). Jazyky přijímané DZA označujeme jako deterministické bezkontextové jazyky ( $\mathcal{L}_{DCF}$ ). Formálně:  $\mathcal{L}_{DCF} \subset \mathcal{L}_2$  a tedy  $\exists L \in \mathcal{L}_2 : L \notin \mathcal{L}_{DCF}$ .

#### Zdroje nedeterminismu NZA

- volba dalšího stavu z množiny stavů ( $q \in 2^{Q \times \Gamma^*}$ ),
- rozhodnutí, zda číst další znak ze vstupu ( $a \in (\Sigma \cup \{\epsilon\})$ ).

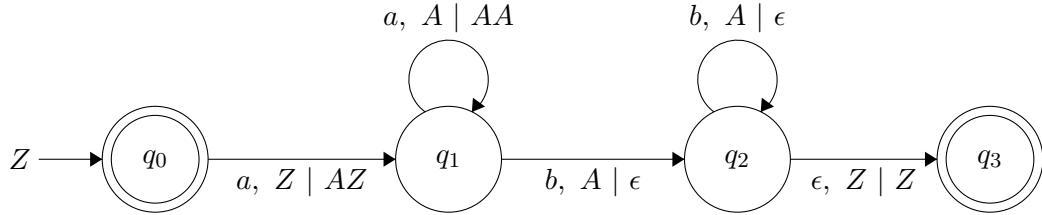
**Definice** Formálně:

$$\forall q \in Q \ \forall a \in \Sigma \ \forall Z \in \Gamma :$$

$$: ( |\delta(q, a, Z)| \leq 1 \wedge |\delta(q, \epsilon, Z)| = 0) \vee$$

$$\vee ( |\delta(q, a, Z)| = 0 \wedge |\delta(q, \epsilon, Z)| \leq 1)$$

**Příklad** Příklad DZA, který přijímá jazyk  $L = \{a^n b^n \mid n \geq 0\}$ . Platí  $L \in \mathcal{L}_{DCF} \setminus \mathcal{L}_3$ .



Obrázek 4.4: Příklad deterministického zásobníkového automatu.

#### 4.3.4 Deterministický rozšířený zásobníkový automat

##### Zdroje nedeterminismu RNZA

- volba dalšího stavu z množiny stavů ( $q \in 2^{Q \times \Gamma^*}$ ),
- rozhodnutí, zda číst další znak ze vstupu ( $a \in (\Sigma \cup \{\epsilon\})$ ),
- rozhodnutí, kolik znaků číst ze zásobníku ( $\gamma \in \Gamma^*$ ).

# Kapitola 5

## TIN – Turingovy stroje (jazyky přijímané TS, varianty TS, lineárně omezené automaty, vyčíslitelné funkce).

### 5.1 Zdroje

- tin\_2021\_merged.pdf
- TIN\_2020-10-20.mp4
- TIN\_2020-10-27.mp4
- TIN\_2020-11-03.mp4
- TIN\_2020-11-24\_p1.mp4

### 5.2 Turingův stroj

- Turingovy stroje jsou velmi robustní a jejich různé úpravy (determinismus, nedeterminismus, počet pásek, ...) mají ekvivalentní vyjadřovací sílu z hlediska rozhodnutelnosti, ale z hlediska složitosti mají různou vyjadřovací sílu.
- Po pásmu se můžeme pohybovat oběma směry. Páska je z prava nekonečná.
- Na TS lze nahlížet také jako na funkci, která má vstup počáteční pásku, pak nějaký výpočet (činnost TS) a jako výstup vrací stav pásky.

**Definice** Turingův stroj (TS) je šestice  $M = (Q, \Sigma, \Gamma, \delta, q_0, g_f)$ , kde

- $Q$  je konečná množina stavů;
- $\Sigma$  je vstupní abeceda (symboly, které se mohou vyskytovat na výchozím stavu pásky),
  - $\Delta \notin \Sigma$ ;
- $\Gamma$  je pásková abeceda (symboly, které je možné zapisovat na pásku),
  - $\Sigma \subset \Gamma \wedge \Delta \in \Gamma \wedge L, R \notin \Gamma$ ;
  - Symbol  $\Delta$  značí tzv. blank (prázdný symbol), který se vyskytuje na místech pásky, která nebyla ještě použita (může ale být na pásku zapsán i později).

- $\delta$  je přechodová funkce (parciální funkce),
  - $\delta : (Q - \{q_f\}) \times \Gamma \rightarrow Q \times (\Gamma \cup \{L, R\});$
- $q_0$  je výchozí stav,
  - $q_0 \in Q;$
- $q_f$  je koncový stav,
  - $q_f \in Q;$

**Konfigurace** Konfigurace TS je trojice  $(q, \alpha, n) \in Q \times \{\gamma\Delta^\omega \mid \gamma \in \Gamma\} \times \mathbb{N}$ , kde

- $q$  je aktuální stav;
- $\alpha$  značí stav pásky;
- $n$  značí pozici hlavy.

**Počáteční konfigurace** Počáteční konfigurace je taková konfigurace  $(q_0, \gamma\Delta^\omega, 0)$ , kde  $\gamma\Delta^\omega$  je výchozí stav pásky a  $q_0$  je výchozí stav.

**Koncová konfigurace** Koncová konfigurace je taková konfigurace  $(q_f, \gamma\Delta^\omega, n)$ , kde  $\gamma\Delta^\omega$  je stav pásky a  $q_f$  je koncový stav.

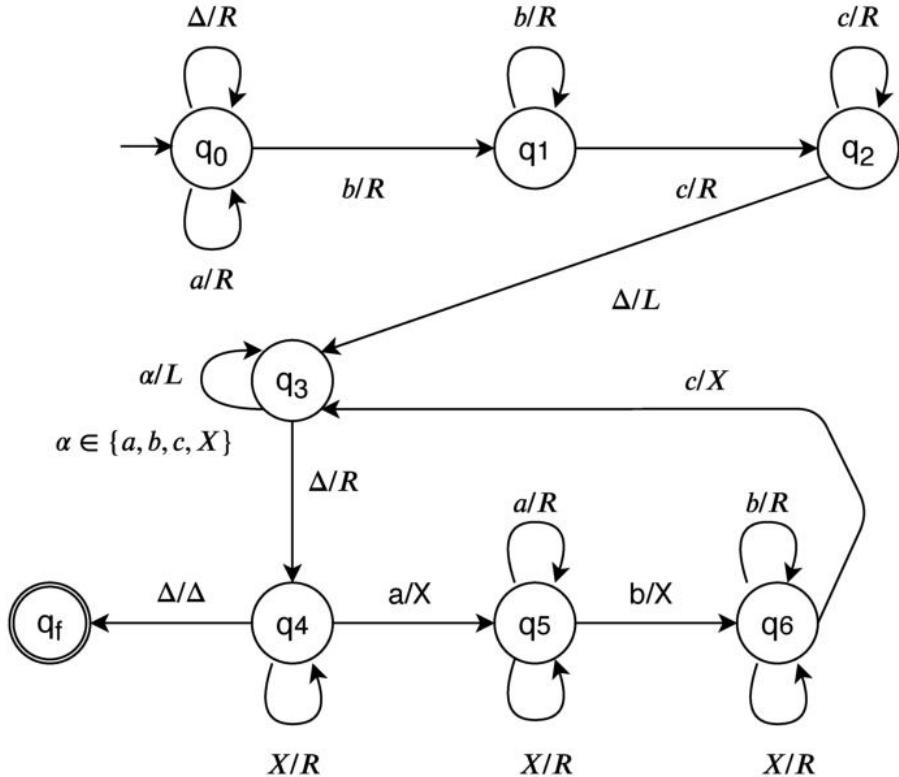
**Přechod** Přechod (krok výpočtu) TS  $M$  definujeme jako nejmenší binární relaci  $\vdash$  takovou, že  $\forall q_1, q_2 \in Q \ \forall \gamma \in \Gamma^\omega \ \forall n \in \mathbb{N} \ \forall b \in \Gamma$ :

- $(q_1, \gamma, n) \vdash (q_2, \gamma, n + 1)$  pro  $\delta(q_1, \gamma_n) = (q_2, R)$  – operace posunu hlavy doprava při  $\gamma_n$  po hlavou;
- $(q_1, \gamma, n) \vdash (q_2, \gamma, n - 1)$  pro  $\delta(q_1, \gamma_n) = (q_2, L)$  – operace posunu hlavy doleva při  $\gamma_n$  po hlavou;
- $(q_1, \gamma, n) \vdash (q_2, s_b^n(\gamma), n + 1)$  pro  $\delta(q_1, \gamma_n) = (q_2, b)$  – operace zápisu symbolu  $b$  při  $\gamma_n$  po hlavou.

Pro libovolný řetězec  $\gamma \in \Gamma^\omega$  a číslo  $n \in \mathbb{N}$  označme  $\gamma_n$   $n$ -tý symbol daného řetězce a označme  $s_n^b(\gamma)$  řetězec, který vznikne z  $\gamma$  záměnou  $\gamma_n$  za  $b$ .

### Jazyk přijímaný

- Řetězec  $w \in \Sigma^*$  je přijat TS  $M = (Q, \Sigma, \Gamma, \delta, q_0, g_f)$ , jesliže pro  $M$  platí:  $(q_0, \Delta\gamma\Delta^\omega, 0) \vdash^* (q_f, \gamma', n)$  pro nějaké  $\gamma, \gamma' \in \Gamma^*$  a  $n \in \mathbb{N}$ .
- Množinu  $L(M) = \{w \mid w \text{ je přijat TS } M\} \subseteq \Sigma^*$  nazýváme jazyk přijímaný TS  $M$ .



Obrázek 5.1: Příklad TS, který přijímá jazyk  $L = \{a^n b^n c^n \mid n > 0\}$ .

**Univerzální turingův stroj** Je možné sestrojit takový turingův stroj, který má vlastnost, že dokáže simulovat chování jiného Turingova stroje. Může fungovat jako interpret. Na vstupu přijme zakódovaný Turingův stroj (program) a vstup, který vykoná.

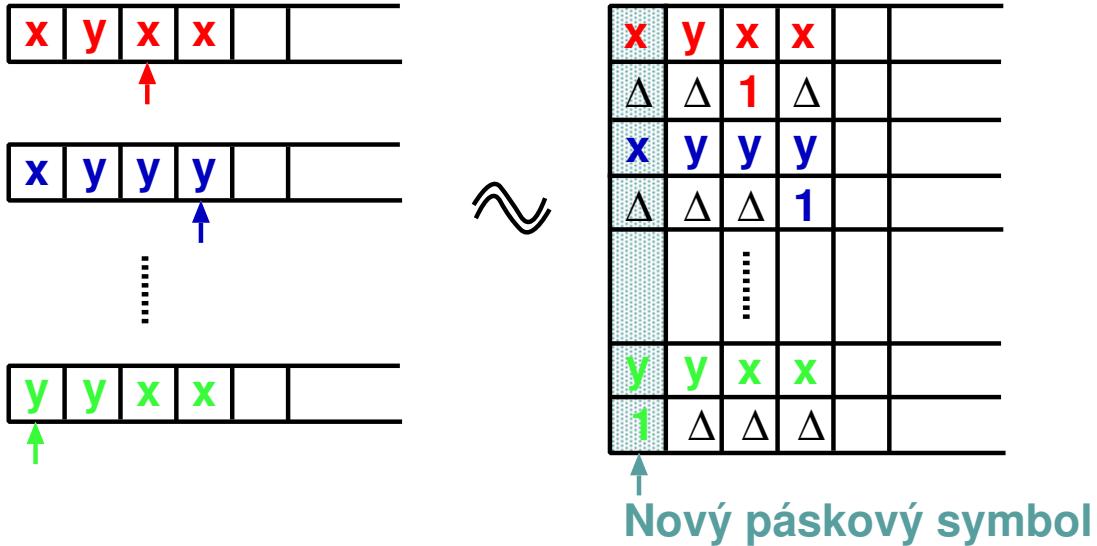
## 5.3 Varianty turingova stroje

### 5.3.1 Deterministický turingův stroj

Deterministický turingův stroj (DTS) je výchozí turingův stroj ( $DTS = TS$ ).

### 5.3.2 Vícepáskový turingův stroj

Vícepáskový TS má stejnou (rozhodovací) vyjadřovací sílu jako jednopáskový TS (jsou mezi sebou převoditelné). Formálně to znamená, že pro každý  $k$ -páskový TS  $M$  existuje jednopáskový TS  $M'$  takový, že  $L(M) = L(M')$ . Myšlenka důkazu spočívá v tom, že  $k$  pásek můžeme simulovat jednou. Jeden symbol na pásmu potom bude mít formu  $2k$ -tice. Pro každou pásku jsou potřeba 2 prvky v  $n$ -tici, protože v jedné je třeba si udržovat pozici hlavy. Myšlenka je na obrázku.



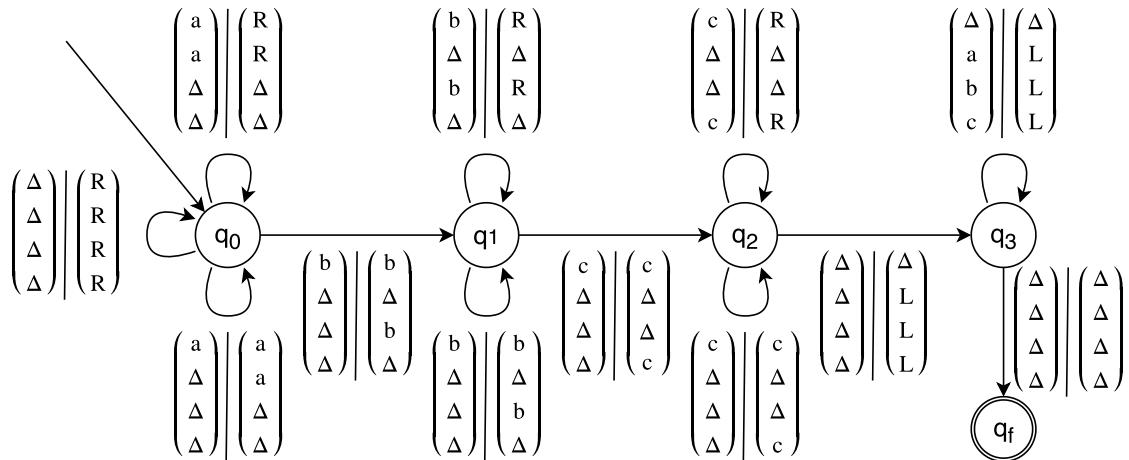
Obrázek 5.2: Myšlenka převodu vícepáskového TS na jednopáskový.

**Definice** Vícepáskový TS se od TS liší tvarem přechodové funkce, disponuje více páskami a s každou může pracovat zároveň. Formálně:

- $\delta$  je přechodová funkce (parciální funkce).

$$\delta : (Q - \{q_f\}) \times \Gamma_1 \times \dots \times \Gamma_n \rightarrow Q \times (\Gamma_1 \cup \{L, R\}) \times \dots \times (\Gamma_n \cup \{L, R\})$$

**Příklad** Příklad vícepáskového TS.



Obrázek 5.3: Příklad 4-páskový TS, který přijímá jazyk  $L = \{a^n b^n c^n \mid n > 0\}$ .

### 5.3.3 Nedeterministický turingův stroj

Nedeterministický TS (NTS) má stejnou (rozhodovací) vyjadřovací sílu jako DTS (jsou mezi sebou převoditelné). Formálně to znamená, že pro každý NTS  $M$  existuje DTS  $M'$  takový, že  $L(M) = L(M')$ . Myšlenka důkazu spočívá v tom, že strom běhů NTS prohledáváme do šířky (jelikož běh může obsahovat smyčky). Konkrétně, NTS  $M$  budeme simulovat třípáskovým DTS. Význam jednotlivých pásek tohoto stroje je následující:

- Páska 1 obsahuje vstupní řetězec.
- Páska 2 je pracovní páska. Obsahuje kopii pásky 1 ohraničenou vhodnými speciálními značkami. Po neúspěšném pokusu o přijetí je její obsah smazán a obnoven z první pásky.
- Páska 3 obsahuje kódovanou volbu posloupnosti přechodů; při neúspěchu bude její obsah nahrazen jinou posloupností.

**Definice** NTS se od TS liší tvarom přechodové funkce. Formálně:

- $\delta$  je přechodová funkce (parciální funkce).

$$\delta : (Q - \{q_f\}) \times \Gamma \rightarrow 2^{Q \times (\Gamma \cup \{L, R\})}$$

**Příklad** Příklad nedeterministického TS, kde nedeterminismus může přinést značné zjednodušení, je TS, který přijímá jazyk  $L = \{ww \mid w \in \Sigma^*\}$ .

### 5.3.4 Úplný turingův stroj

Úplný turingův stroj je takový TS, který pro libovolný vstup zastaví, a bud' přijme nebo odmítne (nemůže se zacyklit). Úplné turingovy stroje mají menší (rozhodovací) vyjadřovací sílu než obecný turingův stroj. Nedeterministický Turingův stroj je úplný, právě když pro každý vstup je každá výpočetní větev konečná (tj. pro každý vstup vždy zastaví).

- Jazyk  $L \subseteq \Sigma^*$  se nazývá:
  - Rekurzivně vyčíslitelný ( $REL, \mathcal{L}_0, \mathcal{L}_{RE}$ ), jestliže  $L = L(M)$  pro nějaký TS  $M$ .
  - Rekurzivní ( $RL, \mathcal{L}_{Rec}$ ), jestliže  $L = L(M)$  pro nějaký úplný TS  $M$ .
- Je-li  $M$  úplný TS, pak říkáme, že  $M$  rozhoduje jazyk  $L(M)$ .
- Ke každému rekurzívnímu jazyku existuje TS, který ho rozhoduje, tj. zastaví pro každé vstupní slovo (např. na pásku zapíše YES nebo NO).
- TS přijímající rekurzivně vyčíslitelný jazyk  $L$  zastaví pro každé  $w \in L$ , ovšem pro  $w \notin L$  může zastavit, ale také může donekonečna cyklist.
- Platí  $2^{\Sigma^*} \subset \mathcal{L}_{RE} \subset \mathcal{L}_{Rec} \subset \mathcal{L}_1$ .

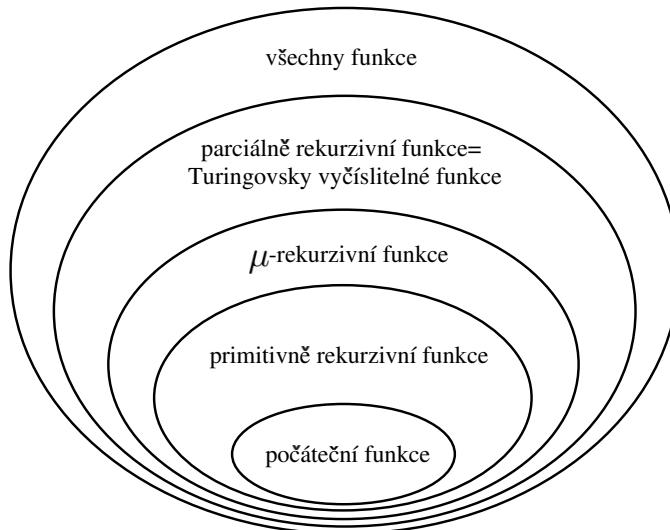
## 5.4 Lineárně omezený automat

- Lineárně omezený automat (LOA) je nedeterministický TS, který nikdy neopustí tu část pásky, na níž je zapsán jeho vstup.
  - Takový turingův stroj, který má lineárně omezenou velikost pásky vůči délce vstupu (*Linear Bounded Automata*). Avšak, díky možnosti zápisu n-tic na pásku (viz důkaz u vícepaskových TS), je možné pásku zkomprimovat tak, že stačí pouze tolik místa, kolik je délka vstupního slova.
- Deterministický LOA můžeme definovat jako DTS, který nikdy neopustí část pásky se zapsaným vstupem.
  - Není známo, zda deterministický LOA je či není striktně slabší než LOA.
- Třída jazyků, kterou lze generovat kontextovými gramatikami, odpovídá třídě jazyků, které lze přijímat LOA.

- Z konečné délky pásky vyplývá, že pro každý LOA existuje pouze konečné množství konfigurací. To znamená, že pokud LOA cyklí, jsme to schopni detektovat. A tedy, problémy pro LOA jsou rozhodnutelné.

## 5.5 Vyčíslitelné funkce

- Tzv. rekurzivní / vyčíslitelné funkce jsou  $f : \mathbb{N}^n \rightarrow \mathbb{N}^m$ , kde  $n, m \in \mathbb{N}$ .
- Klasifikace parciálních funkcí:
  - Totální funkce (např. funkce sčítání) – funkce je definovaná pro celý definiční obor, tj. pro každý vstup existuje výstup.
  - Striktně parciální funkce (např. funkce dělení) – funkce není definovaná pro celý definiční obor, tj. existuje vstup jemuž není definovaný výstup.
- Funkce totální, nemusí být vyčíslitelná (není možné je vypočítat počítačem, ačkoliv matematicky je lze zadefinovat).
- Konvence:  $n$ -tici  $(x_1, x_2, \dots, x_n) \in \mathbb{N}^n$  budeme označovat jako  $\bar{x}$ .



Obrázek 5.4: Hierarchie vyčíslitelných funkcí.

### 5.5.1 Počáteční funkce

Základní funkce, nacházejí se nejnižše v hierarchii vyčíslitelných funkcí.

- Nulová funkce  $[\xi]$  (*zero function*)

$$\xi : () \rightarrow 0$$

$$\xi() = 0$$

- Funkce následníka  $[\sigma]$  (*successor function*)

$$\sigma : \mathbb{N} \rightarrow \mathbb{N}$$

$$\sigma(x) = x + 1$$

- Projekce  $[\pi]$  (*projection*) – Vybírá z  $n$ -tice  $k$ -tý prvek.

$$\pi_k^n : \mathbb{N}^n \rightarrow \mathbb{N}$$

$$\pi_k^n(\bar{x}) = x_k$$

### 5.5.2 Primitivně rekurzivní funkce

Jde o funkce vytvořené z počátečních funkcí.

- Kombinace  $[\times]$  – Kombinace dvou funkcí  $f : \mathbb{N}^k \rightarrow \mathbb{N}^n$  a  $g : \mathbb{N}^k \rightarrow \mathbb{N}^m$  vznikne funkce:

$$f \times g : \mathbb{N}^k \rightarrow \mathbb{N}^{n+m}$$

$$f \times g(\bar{x}) = (f(\bar{x}), g(\bar{x})) , \bar{x} \in \mathbb{N}^k$$

- Kompozice  $[\circ]$  – Kompozice dvou funkcí  $f : \mathbb{N}^k \rightarrow \mathbb{N}^n$  a  $g : \mathbb{N}^n \rightarrow \mathbb{N}^m$  vznikne funkce:

$$g \circ f : \mathbb{N}^k \rightarrow \mathbb{N}^m$$

$$g \circ f(\bar{x}) = g(f(\bar{x})) , \bar{x} \in \mathbb{N}^k$$

- Primitivní rekurze – Primitivní rekurze je technika, která umožňuje vytvořit funkci  $f : \mathbb{N}^{k+1} \rightarrow \mathbb{N}^m$  na základě jiných dvou funkcí  $g : \mathbb{N}^k \rightarrow \mathbb{N}^m$  a  $h : \mathbb{N}^{k+m+1} \rightarrow \mathbb{N}^m$  rovnicemi:

$$f(\bar{x}, 0) = g(\bar{x})$$

$$f(\bar{x}, y + 1) = h(\bar{x}, y, f(\bar{x}, y)) , \bar{x} \in \mathbb{N}^k$$

Následují příklady definic funkcí, které spadají do třídy primitivně rekurzivních funkcí.

**Konstantní funkce** Funkce  $\kappa_m^n$  libovolné  $n$ -tici  $x \in \mathbb{N}^n$  přiřadí konstantní hodnotu  $m \in \mathbb{N}$ .

$$\kappa_m^0 \equiv \sigma \circ \sigma \circ \dots \circ \sigma \circ \xi$$

( $\sigma$  je aplikována  $m$ -krát)

#### Funkce sčítání

$$plus : \mathbb{N}^2 \rightarrow \mathbb{N}$$

$$plus(x, 0) = \pi_1^1(x)$$

$$plus(x, y + 1) = \sigma \circ plus(x, y)$$

#### Funkce násobení

$$mult : \mathbb{N}^2 \rightarrow \mathbb{N}$$

$$mult(x, 0) = \xi()$$

$$mult(x, y + 1) = plus(x, mult(x, y))$$

#### Funkce umocňování

$$exp : \mathbb{N}^2 \rightarrow \mathbb{N}$$

$$exp(x, 0) = \sigma \circ \xi()$$

$$exp(x, y + 1) = mult(x, exp(x, y))$$

### Funkce předchůdce

$$pred : \mathbb{N} \rightarrow \mathbb{N}$$

$$pred(0) = \xi()$$

$$pred(x + 1) = \pi_1^2(x, pred(x))$$

**Funkce monus** Funkce odčítání, ale nejmenší číslo je 0.

$$monus : \mathbb{N}^2 \rightarrow \mathbb{N}$$

$$monus(x, 0) = \pi_1^1(x)$$

$$monus(x, y + 1) = pred \circ monus(x, y)$$

**Funkce negace** Logická negace, 0 je *false*, ostatní čísla jsou *true*.

$$neg : \mathbb{N} \rightarrow \mathbb{N}$$

$$neg(x) = monus(\sigma \circ \xi(), x)$$

**Funkce rovnost** Vrací 0 – *false* a 1 – *true*.

$$eq : \mathbb{N}^2 \rightarrow \mathbb{N}$$

$$eq \equiv monus \circ (\kappa_1^2 \times (plus \circ ((monus \circ (\pi_2^2 \times \pi_1^2)) \times monus \circ (\pi_1^2 \times \pi_2^2))))$$

**Funkce je větší** Vrací 0 – *false* a 1 – *true*.

$$isGreater : \mathbb{N}^2 \rightarrow \mathbb{N}$$

$$isGreater(x, y) = neg \circ eq(monus(x, y), \xi())$$

**Funkce odmocnina** Funkce *foo* je pomocná.

$$foo : \mathbb{N}^2 \rightarrow \mathbb{N}$$

$$foo(0, y) = \xi()$$

$$foo(x + 1, y) = plus(isGreater(mult(\sigma \times \sigma(x)), y), foo(x, y))$$

$$sqrt : \mathbb{N} \rightarrow \mathbb{N}$$

$$sqrt(x) = monus(x, foo(x, x))$$

### 5.5.3 Parciálně rekurzivní funkce

- Zavádí koncept tzv. *minimalizace*.
- Umožňuje realizaci *while* cyklu.
- Parciálně rekurzivní funkce jsou ekvivalentní Turingovým strojům.
- **[[todo: dokončit]]**

#### 5.5.4 Turingovsky vyčíslitelné funkce

- Parciální funkce, kterou může počítat nějaký turingův stroj se nazývá turingovsky vyčíslitelná funkce.
- Každá parciálně rekurzivní funkce je turingovsky vyčíslitelná.
- Každý výpočetní proces prováděný turingovým strojem je procesem vyčíslení nějaké parciálně rekurzivní funkce.

**Definice** Turingův stroj  $M = (Q, \Sigma, \Gamma, \delta, q_0, g_f)$  vyčísluje (počítá) parciální funkci  $f : (\Sigma^*)^m \rightarrow (\Sigma_1^*)^n$ ,  $\Sigma_1 \subseteq \Gamma$ ,  $\Delta \notin \Sigma_1$ , jestliže pro každé  $(w_1, w_2, \dots, w_m) \in (\Sigma^*)^m$  a odpovídající počáteční konfiguraci  $\Delta w_1 \Delta w_2 \Delta \dots \Delta w_m \Delta^\omega$  TS  $M$ :

1. V případě, že  $f(w_1, w_2, \dots, w_m)$  je definována, pak  $M$  zastaví a páška obsahuje  $\Delta v_1 \Delta v_2 \Delta \dots \Delta v_n \Delta^\omega$ , kde  $(v_1, v_2, \dots, v_n) = f(w_1, w_2, \dots, w_m)$ .
2. V případě, že  $f(w_1, w_2, \dots, w_m)$  není definována,  $M$  cyklí nebo zastaví abnormálně.

# Kapitola 6

## TIN – Nerozhodnutelnost (problém zastavení TS, princip diagonalizace a redukce, Postův korespondenční problém).

### 6.1 Zdroje

- tin\_2021\_merged.pdf
- TIN\_2020-11-03.mp4
- TIN\_2020-11-10.mp4
- TIN\_2020-11-20\_demo.mp4

### 6.2 Rozhodovací problém

#### Rozhodovací problém

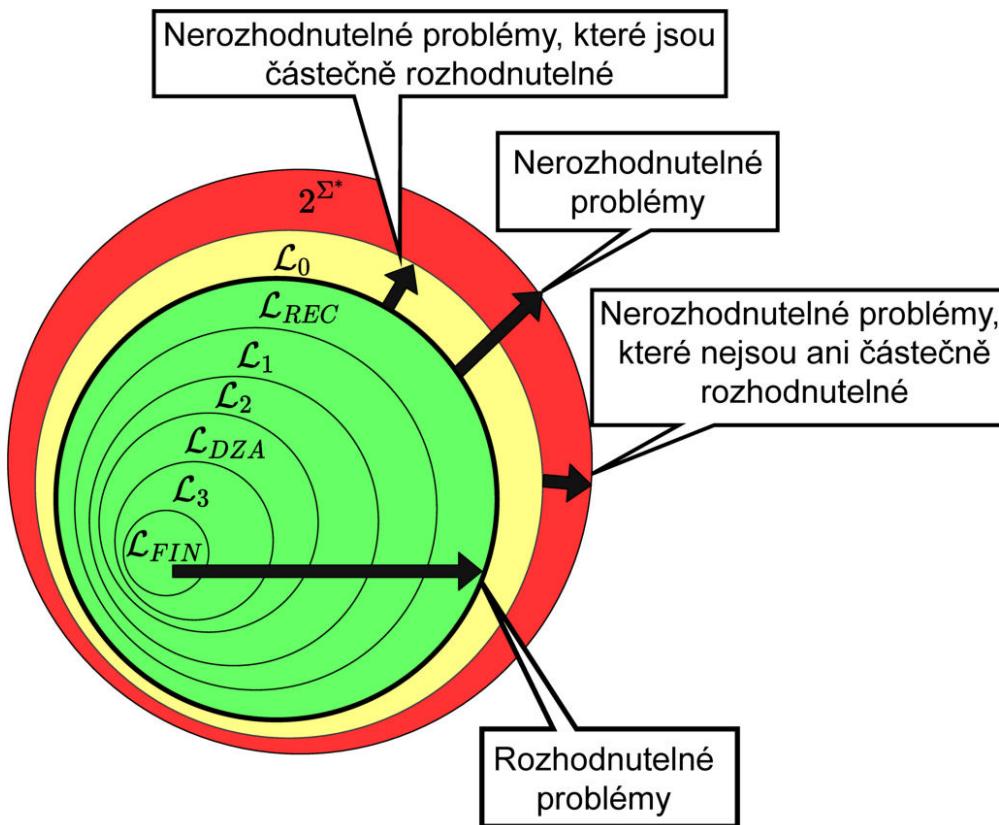
- Rozhodovací problém (*decision problem*)  $P$  může být chápán jako funkce  $f_P$  s oborem hodnot  $\{true, false\}$ .
- Rozhodovací problém je obvykle specifikován:
  - definičním oborem  $A_P$  reprezentujícím množinu možných instancí problému,
  - podmnožinou  $B_P \subseteq A_P$ ,  $B_P = \{p \mid f_P(p) = true\}$  instancí, pro které je hodnota  $f_P$  rovna *true*.

#### Kódování problémů

- V teorii formálních jazyků používáme ke kódování jednotlivých instancí problémů řetězce nad vhodnou abecedou  $\Sigma$ .
- Pak je rozhodovací problém  $P$  přirozeně specifikován jazykem  $L_P = \{w \in \Sigma^* \mid w = code(p), p \in B_P\}$ , kde  $code : A_P \rightarrow \Sigma^*$  je injektivní funkce, která přiřazuje instancím problému příslušný řetězec (nezávisle na  $f_P$ ).

#### Rozhodování problémů TS

- Nechť  $P$  je problém specifikovaný jazykem  $L_P$  nad abecedou  $\Sigma$ . Problém  $P$  nazveme:
  - **Rozhodnutelný**, pokud  $L_P$  je rekurzívny jazyk, tj. existuje TS, ktorý  $L_P$  rozhoduje (prijme každý řetězec  $w \in L_P$ , a zamíne každý řetězec  $w \in \Sigma^* - L_P$ ).
  - **Nerozhodnutelný**, když není rozhodnutelný.
  - **Částečně rozhodnutelný**, jestliže  $L_P$  je rekurzívne vyčísliteľny jazyk.
- Z toho plynne, že každý rozhodnutelný problém je současně částečně rozhodnutelný, ale některé nerozhodnutelné problémy nejsou ani částečně rozhodnutelné.
  - Problém je nerozhodnutelný, pokud pro něj neexistuje úplný turingov stroj.



Obrázek 6.1: Hierarchie jazýků/problémů a jejich (ne)rozhodnutelnost.

## 6.3 Známé problémy

### Problém zastavení TS

- Problém zastavení TS  $M$  na daném vstupu  $w$  (HP, *Halting Problem*).
- Neformálně: Znáte-li zdrojový kód programu (TS) a jeho vstup, rozhodněte, zda program zastaví, nebo zda poběží navždy bez zastavení.
- Problém není rozhodnutelný, ale je částečně rozhodnutelný.
  - $HP \in \mathcal{L}_{RE}$
  - Důkaz: s využitím Cantorovy diagonalizace.

$$HP = \{\langle M \rangle \# \langle w \rangle \mid \text{TS } M \text{ na vstupu } w \text{ zastaví}\}$$

### Problém nezastavení TS

- Problém nezastavení TS  $M$  na daném vstupu  $w$  (co-HP, *co-Halting Problem*).
- Neformálně: Znáte-li zdrojový kód programu (TS) a jeho vstup, rozhodněte, zda program nezastaví (poběží navždy bez zastavení), nebo zda zastaví.
- Problém není ani částečně rozhodnutelný.
  - $\text{coHP} \notin \mathcal{L}_{RE}$
  - Důkaz: vyplývá z definice komplementu.

$$\text{coHP} = \{\langle M \rangle \# \langle w \rangle \mid \text{TS } M \text{ na vstupu } w \text{ nezastaví}\}$$

### Problém náležitosti

- Problém náležitosti řetězce  $w$  do jazyka  $L$ ,  $L \in \mathcal{L}_0$  (MP, *Membership Problem*).
- Neformálně: Znáte-li TS a jeho vstup, rozhodněte, zda TS daný vstup přijme (zastaví a akceptuje) a nebo nepřijme (zastaví a neakceptuje, zastaví abnormálně, nezastaví).
  - Jde o podobný případ jako problém zastavení TS, pouze obsahuje krok navíc.
- Problém není rozhodnutelný, ale je částečně rozhodnutelný.
  - $MP \in \mathcal{L}_{RE}$
  - Důkaz: s využitím redukce z HP.

$$MP = \{\langle M \rangle \# \langle w \rangle \mid w \in L(M)\}$$

### Problém nenáležitosti

- Problém nenáležitosti řetězce  $w$  do jazyka  $L$ ,  $L \in \mathcal{L}_0$  (coMP, *Non-Membership Problem*).
- Neformálně: Znáte-li TS a jeho vstup, rozhodněte, zda TS daný vstup nepřijme (zastaví a neakceptuje, zastaví abnormálně, nezastaví) a nebo přijme (zastaví a akceptuje).
- Problém není ani částečně rozhodnutelný.
  - $\text{coMP} \notin \mathcal{L}_{RE}$
  - Důkaz: s využitím redukce z coHP.

$$\text{coMP} = \{\langle M \rangle \# \langle w \rangle \mid w \notin L(M)\}$$

### Postův korespondenční problém

- Postův systém  $S$  nad abecedou  $\Sigma$  je dán neprázdným seznamem dvojic neprázdných řetězců nad  $\Sigma$ , formálně:

$$S = \langle (\alpha_1, \beta_1), \dots, (\alpha_k, \beta_k) \rangle, \text{ pro } k \geq 1, \forall i : 1 \leq i \leq k : \alpha_i, \beta_i \in \Sigma^+$$

- Řešením Postova systému  $S$  je každá neprázdná posloupnost přirozených čísel  $I = \langle i_1, i_2, \dots, i_m \rangle$ , kde  $m \geq 1$  a  $1 \leq j \leq m : 1 \leq i_j \leq k$ , tak, že:

$$\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_m} = \beta_{i_1}, \beta_{i_2}, \dots, \beta_{i_m}$$

- $m$  není omezené a indexy se mohou opakovat.
- Postův korespondenční problém (PCP) zní: existuje pro daný Postův systém řešení?
- Problém není ani částečně rozhodnutelný.
  - $PCP \notin \mathcal{L}_{RE}$
  - Důkaz: *nezkouší se*
- Příklad: PS  $S$  nad  $\Sigma = \{a, b\}$ :
  - $S = \langle (ab, a), (aa, baaab), (aa, a) \rangle$
  - $S$  má řešení, např.  $I = \langle 1, 2, 1, 3 \rangle$
  - $* ab.aa.ab.aa = a.baaab.a.a$

## Další problémy

- $\{\langle M \rangle \mid TS M \text{ má aspoň 2020 stavů}\} \in \mathcal{L}_{Rec}$
- $\{\langle M \rangle \mid TS M \text{ učiní více než 2020 kroků na vstupu } \epsilon\} \in \mathcal{L}_{Rec}$
- $\{\langle M \rangle \mid TS M \text{ učiní více než 2020 kroků na nějakém vstupu}\} \in \mathcal{L}_{Rec}$
- $\{\langle M \rangle \mid L(M) \neq \emptyset\} \in \mathcal{L}_{RE}$
- $\{\langle M \rangle \mid |L(M)| \geq 42\} \in \mathcal{L}_{RE}$
- $\{\langle M \rangle \mid L(M) = \emptyset\} \notin \mathcal{L}_{RE}$
- $\{\langle M \rangle \mid |L(M)| \leq 42\} \notin \mathcal{L}_{RE}$
- $\{\langle M \rangle \mid L(M) \in \mathcal{L}_{Rec}\} \notin \mathcal{L}_{RE}$

## 6.4 Diagonalizace

- Diagonalizace (Cantorova diagonální metoda) je důkázová technika sporem, která slouží k porovnání mohutnosti dvou množin (zejména dvou nekonečných množin).
- Spočetné nekonečno:
  - Dokážeme „spočítat“ a usporádat.
  - Existuje bijekce mezi množinou  $\mathbb{N}$  a každou spočetně nekonečnou množinou.
- Nespočetné nekonečno:
  - Nedokážeme „spočítat“ a usporádat.
  - Neexistuje bijekce mezi množinou  $\mathbb{N}$  a žádnou nespočetně nekonečnou množinou.

### Příklad 1: dokažte, že $\mathbb{R}$ je větší než $\mathbb{N}$

- Zřejmě platí, že pokud podmnožiny množiny je nespočetná, tak i množina je nespočetná.
  - Z toho vyplývá, že stačí dokázat že libovolná podmnožina  $M \subseteq \mathbb{R}$  je nespočetná.

- Vybereme množinu  $M = (0, 1)$ .
- Předpokládejme, že  $M$  je spočetná množina, pak existuje bijekce  $f : \mathbb{N} \leftrightarrow M$ .
- Uspořádejme  $M$  do libovolné posloupnosti a zobrazme  $f$  jako nekonečnou matici.

|        | 1. desetinná pozice | 2. desetinná pozice | 3. desetinná pozice | ... |
|--------|---------------------|---------------------|---------------------|-----|
| $f(0)$ | $d_{11}$            | $d_{12}$            | $d_{13}$            | ... |
| $f(1)$ | $d_{21}$            | $d_{22}$            | $d_{23}$            | ... |
| $f(2)$ | $d_{31}$            | $d_{32}$            | $d_{33}$            | ... |
| ...    | ...                 | ...                 | ...                 | ... |

Kde  $\forall i, j : d_{ij} \in \{0, 1, \dots, 9\}$  a  $\forall n \in \mathbb{N} : f(n) = 0, d_{n1}d_{n2} \dots$

- Uvažme číslo  $x = 0, e_1e_2e_3 \dots$  takové, že  $e_i = 9 - d_{ii}$  pro  $i = 1, 2, 3, \dots$ .
- Zřejmě platí, že  $x \in M$  a tedy, musí existovat  $n \in \mathbb{N} : f(n) = x$ .
- Ale současně  $\forall n \in \mathbb{N} : f(n) \neq x$  protože dojde k neshodě minimálně na diagonále.
  - Spor.

**Příklad 2:** dokažte, že  $2^{\Sigma^*}$  je větší než  $\mathcal{L}_0$

**Věta 9.1** Pro každou abecedu  $\Sigma$  existuje jazyk nad  $\Sigma$ , který není typu 0 (tj. rekurzivně vyčíslitelný).

*Důkaz.*

1. Libovolný jazyk typu 0 nad  $\Sigma$  může být přijat TS s  $\Gamma = \Sigma \cup \{\Delta\}$ : Pokud  $M$  používá více symbolů, můžeme je zakódovat jako jisté posloupnosti symbolů ze  $\Sigma \cup \{\Delta\}$  a sestrojit TS  $M'$ , který simuluje  $M$ .
2. Nyní můžeme snadno systematicky vypisovat všechny TS s  $\Gamma = \Sigma \cup \{\Delta\}$ .  
Začneme stroji se dvěma stavami, pak se třemi stavami, ...  
**Závěr:** Množina všech takových strojů a tedy i jazyků typu 0 je spočetná.
3. Množina  $\Sigma^*$  ale obsahuje nekonečně mnoho řetězců a proto je množina  $2^{\Sigma^*}$  zahrnující všechny jazyky nespočetná – důkaz viz další strana.
4. Z rozdílnosti mohutností spočetných a nespočetných množin plyne platnost uvedené věty.

□

**Lemma 9.1** Pro neprázdné, konečné  $\Sigma$  je množina  $2^{\Sigma^*}$  nespočetná.

*Důkaz.* Důkaz provedeme tzv. **diagonalizací** (poprvé použitou Cantorem při důkazu rozdílné mohutnosti  $\mathbb{N}$  a  $\mathbb{R}$ ).

- Předpokládejme, že  $2^{\Sigma^*}$  je spočetná. Pak dle definice spočetnosti existuje **bijekce**  $f : \mathbb{N} \longleftrightarrow 2^{\Sigma^*}$ .
- Uspořádejme  $\Sigma^*$  do nějaké posloupnosti  $w_1, w_2, w_3, \dots$ , např.  $\varepsilon, x, y, xx, xy, yx, yy, xxx, \dots$  pro  $\Sigma = \{x, y\}$ . Nyní můžeme  $f$  zobrazit **nekonečnou maticí**:

|              | $w_0$    | $w_1$    | $w_2$    | ... | $w_i$    | ... |
|--------------|----------|----------|----------|-----|----------|-----|
| $L_0 = f(0)$ | $a_{00}$ | $a_{01}$ | $a_{02}$ | ... | $a_{0i}$ | ... |
| $L_1 = f(1)$ | $a_{10}$ | $a_{11}$ | $a_{12}$ | ... | $a_{1i}$ | ... |
| $L_2 = f(2)$ | $a_{20}$ | $a_{21}$ | $a_{22}$ | ... | $a_{2i}$ | ... |
| ...          | ...      | ...      | ...      | ... | ...      | ... |

- Uvažujme jazyk  $\bar{L} = \{w_i \mid a_{ii} = 0\}$ .  $\bar{L}$  se liší od každého jazyka  $L_i = f(i)$ ,  $i \in \mathbb{N}$ :
  - je-li  $a_{ii} = 0$ , pak  $w_i$  patří do jazyka,
  - je-li  $a_{ii} = 1$ , pak  $w_i$  nepatří do jazyka.
- Současně ale  $\bar{L} \in 2^{\Sigma^*}$ ,  $f$  tudíž není surjektivní, což je spor.

□

**Příklad 3:** dokažte, že HP není rozhodnutelný, ale je částečně rozhodnutelný

**Věta 9.2 Problém zastavení TS (Halting Problem)**, kdy nás zajímá, zda daný TS  $M$  pro danou vstupní větu  $w$  zastaví, **není rozhodnutelný**, ale je **částečně rozhodnutelný**.

*Důkaz.*

- Problému zastavení odpovídá rozhodování jazyka  $HP = \{\langle M \rangle \# \langle w \rangle \mid M \text{ zastaví při } w\}$ , kde  $\langle M \rangle$  je kód TS  $M$  a  $\langle w \rangle$  je kód  $w$ .
- **Částečnou rozhodnutelnost** ukážeme snadno použitím modifikovaného univerzálního TS  $T_U$ , který zastaví přijetím vstupu  $\langle M \rangle \# \langle w \rangle$  právě tehdy, když  $M$  zastaví při  $w$  – modifikace spočívá v převedení abnormálního zastavení při simulaci na zastavení přechodem do  $q_F$ .
- **Nerozhodnutelnost** ukážeme pomocí **diagonalizace**:
  1. Pro  $x \in \{0, 1\}^*$ , nechť  $M_x$  je TS s kódem  $x$ , je-li  $x$  legální kód TS. Jinak ztotožníme  $M_x$  s pevně zvoleným TS, např. TS, který pro libovolný vstup okamžitě zastaví.
  2. Můžeme nyní sestavit posloupnost  $M_\varepsilon, M_0, M_1, M_{00}, M_{01}, M_{10}, M_{11}, M_{000}, \dots$  zahrnující všechny TS nad  $\Sigma = \{0, 1\}$  indexované řetězci z  $\{0, 1\}^*$ .

3. Uvažme nekonečnou matici

|                 | $\varepsilon$                    | 0                      | 1                      | 00                      | 01                      | 10                      | ... |
|-----------------|----------------------------------|------------------------|------------------------|-------------------------|-------------------------|-------------------------|-----|
| $M_\varepsilon$ | $H_{M_\varepsilon, \varepsilon}$ | $H_{M_\varepsilon, 0}$ | $H_{M_\varepsilon, 1}$ | $H_{M_\varepsilon, 00}$ | $H_{M_\varepsilon, 01}$ | $H_{M_\varepsilon, 10}$ | ... |
| $M_0$           | $H_{M_0, \varepsilon}$           | $H_{M_0, 0}$           | $H_{M_0, 1}$           | $H_{M_0, 00}$           | $H_{M_0, 01}$           | $H_{M_0, 10}$           | ... |
| $M_1$           | $H_{M_1, \varepsilon}$           | $H_{M_1, 0}$           | $H_{M_1, 1}$           | $H_{M_1, 00}$           | $H_{M_1, 01}$           | $H_{M_1, 10}$           | ... |
| $M_{00}$        | $H_{M_{00}, \varepsilon}$        | $H_{M_{00}, 0}$        | $H_{M_{00}, 1}$        | $H_{M_{00}, 00}$        | $H_{M_{00}, 01}$        | $H_{M_{00}, 10}$        | ... |
| $M_{01}$        | $H_{M_{01}, \varepsilon}$        | $H_{M_{01}, 0}$        | $H_{M_{01}, 1}$        | $H_{M_{01}, 00}$        | $H_{M_{01}, 01}$        | $H_{M_{01}, 10}$        | ... |
| ...             |                                  |                        |                        |                         |                         |                         |     |

kde  $H_{M_x, y} = \begin{cases} \mathbf{C}, & \text{jestliže } M_x \text{ cyklí na } y, \\ \mathbf{Z}, & \text{jestliže } M_x \text{ zastaví na } y. \end{cases}$

4. Předpokládejme, že existuje úplný TS  $K$  přijímající jazyk  $HP$ , tj.  $K$  pro vstup  $\langle M \rangle \# \langle w \rangle$

- zastaví normálně (přijme) právě tehdy, když  $M$  zastaví na  $w$ ,
- zastaví abnormálně (odmítne) právě tehdy, když  $M$  cyklí na  $w$ .

5. Sestavíme TS  $N$ , který pro vstup  $x \in \{0, 1\}^*$ :

- Sestaví  $M_x$  z  $x$  a zapíše  $\langle M_x \rangle \# x$  na svou pásku.
- Simuluje  $K$  na  $\langle M_x \rangle \# x$ , přijme, pokud  $K$  odmítne, a přejde do nekonečného cyklu, pokud  $K$  přijme.

Všimněme si, že  $N$  v podstatě komplementuje diagonálu uvedené matice:

|                 | $\varepsilon$                    | 0                      | 1                      | 00                      | 01                      | 10                      | ... |
|-----------------|----------------------------------|------------------------|------------------------|-------------------------|-------------------------|-------------------------|-----|
| $M_\varepsilon$ | $H_{M_\varepsilon, \varepsilon}$ | $H_{M_\varepsilon, 0}$ | $H_{M_\varepsilon, 1}$ | $H_{M_\varepsilon, 00}$ | $H_{M_\varepsilon, 01}$ | $H_{M_\varepsilon, 10}$ | ... |
| $M_0$           | $H_{M_0, \varepsilon}$           | $H_{M_0, 0}$           | $H_{M_0, 1}$           | $H_{M_0, 00}$           | $H_{M_0, 01}$           | $H_{M_0, 10}$           | ... |
| $M_1$           | $H_{M_1, \varepsilon}$           | $H_{M_1, 0}$           | $H_{M_1, 1}$           | $H_{M_1, 00}$           | $H_{M_1, 01}$           | $H_{M_1, 10}$           | ... |
| $M_{00}$        | $H_{M_{00}, \varepsilon}$        | $H_{M_{00}, 0}$        | $H_{M_{00}, 1}$        | $H_{M_{00}, 00}$        | $H_{M_{00}, 01}$        | $H_{M_{00}, 10}$        | ... |
| $M_{01}$        | $H_{M_{01}, \varepsilon}$        | $H_{M_{01}, 0}$        | $H_{M_{01}, 1}$        | $H_{M_{01}, 00}$        | $H_{M_{01}, 01}$        | $H_{M_{01}, 10}$        | ... |
| ...             |                                  |                        |                        |                         |                         |                         |     |

6. Dostáváme, že

$$\begin{aligned} N \text{ zastaví na } x &\Leftrightarrow K \text{ odmítne } \langle M_x \rangle \# \langle x \rangle \quad (\text{definice } N) \\ &\Leftrightarrow M_x \text{ cyklí na } x \quad (\text{předpoklad o } K). \end{aligned}$$

7. To ale znamená, že  $N$  se liší od každého  $M_x$  alespoň na jednom řetězci – konkrétně  $x$ . Což je ovšem spor s tím, že posloupnost

$M_\varepsilon, M_0, M_1, M_{00}, M_{01}, M_{10}, M_{11}, M_{000}, \dots$  zahrnuje všechny TS nad  $\Sigma = \{0, 1\}$ .

Tento spor plyne z předpokladu, že existuje TS  $K$ , který pro daný TS  $M$  a daný vstup  $x$  určí (rozhodne), zda  $M$  zastaví na  $x$ , či nikoliv.

□

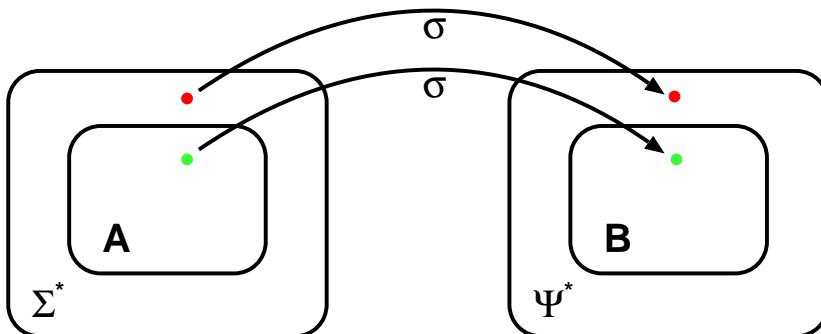
- ❖ Ukázali jsme, že problém zastavení TS je částečně rozhodnutelný a tedy jazyk  $HP$  rekurzívne vyčíslitelný. Z věty 8.6 pak plyne, že **komplement problému zastavení není ani částečně rozhodnutelný** a **jazyk co- $HP = \{\langle M \rangle \# \langle w \rangle \mid M \text{ nezastaví při } w\}$**  je příkladem jazyka, jenž není ani rekurzívne vyčíslitelný.

## 6.5 Redukce

- Redukce je důkazová technika, která slouží k dokázání, že nějaký problém není rozhodnutelný (částečně rozhodnutelný) – neboli, že určitý jazyk není rekurzívni (rekurzívne vyčíslitelný).
  - Víme, že jazyk  $A$  není rekurzívni (rekurzívne vyčíslitelný) a zkoumáme jazyk  $B$ . Ukážeme, že  $A$  lze úplným TS převést (redukovat) na  $B$ , to ale znamená, že  $B$  rovněž není rekurzívni (rekurzívne vyčíslitelný) – jinak by šlo použít úplný TS (ne-úplný TS) přijímající  $B$  a příslušné redukce k sestavení úplného TS (ne-úplného TS) přijímajícího  $A$ , což by byl spor.
- Redukce vytváří uspořádání na problémech (nějaký problém je aspoň tak těžký jako jiný problém).

**Definice** Mějme jazyky  $L_1 \subseteq \Sigma_1^*$  a  $L_2 \subseteq \Sigma_2^*$ . Redukce jazyka  $L_1$  na jazyk  $L_2$  je funkce  $\sigma : \Sigma_1^* \rightarrow \Sigma_2^*$  taková, že

- $\sigma$  je implementovatelná úplným TS (totální rekurzívne vyčíslitelná funkce),
- $\forall w \in \Sigma_1^* : w \in L_1 \Leftrightarrow \sigma(w) \in L_2$  (zachovává členství v jazyce).



Redukce  $\sigma : \Sigma^* \rightarrow \Psi^*$ .

### Použití

- 1. směr:
  - Máme známý problém  $P_1$ , o kterém víme, že je (částečně) rozhodnutelný. To znamená, že existuje TS, který ho (částečně) rozhoduje.
  - Máme nový problém  $P_2$ , o kterém chceme dokázat, že je (částečně) rozhodnutelný. Redukujeme  $P_2$  na  $P_1$ .
  - Cílem je najít takovou funkci  $\sigma$  (překladač), která namapuje instance problému  $P_2$  na instance problému  $P_1$ , takovým způsobem, že není porušena definice redukce.

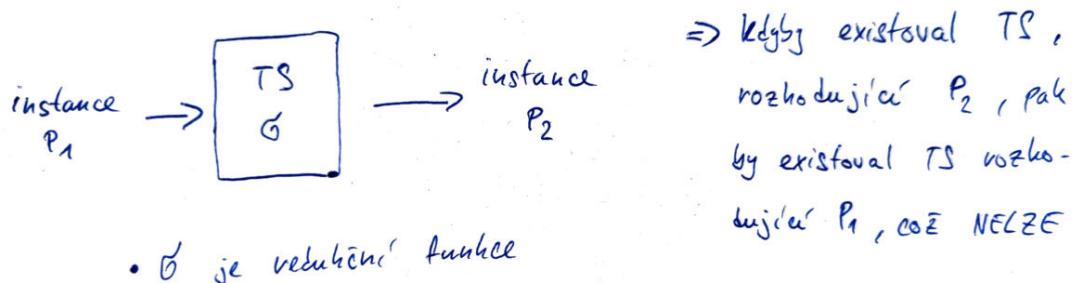
- Z toho vyplývá, že je-li  $P_1$  (částečně) rozhodnutelný, tak i  $P_2$  je (částečně) rozhodnutelný.



•  $M$  rozkazuje  $P_1$       •  $G$  je redukční funkce

- 2. směr:

- Máme známý problém  $P_1$ , o kterém víme, že je nerozhodnutelný (ani částečně).
- Máme nový problém  $P_2$ , o kterém chceme dokázat, že je nerozhodnutelný (ani částečně).
- Cílem je najít takovou funkci  $\sigma$  (překladač), která namapuje instance problému  $P_1$  na instance problému  $P_2$ , takovým způsobem, že není poručena definice redukce.
- Z toho vyplývá, že je-li  $P_1$  nerozhodnutelný (ani částečně), tak i  $P_2$  je nerozhodnutelný (ani částečně).



### Obecný postup důkazu redukcí

1. Navrhnut redukci.
2. Ukázat, že redukce lze implementovat úplným  $TS$ .
3. Ukázat, zachování členství v jazyce.

### Příklad 1: dokažte, že problém neprázdnosti jazyka $TS$ je nerozhodnutelný

- Dokážeme redukcí z problému zastavení  $TS$  (HP).
- Jazyk charakterizující problém zastavení (obsahuje takové instance, pro které  $TS M$  na  $w$  zastaví):

$$HP = \{ \langle M \rangle \# \langle w \rangle \mid TS M \text{ na vstupu } w \text{ zastaví} \}$$

- Jazyk charakterizující problém neprázdnosti (obsahuje takové instance, pro které jazyk TS  $M$  není prázdný):

$$NEMP = \{\langle M \rangle \mid M \text{ je TS takový, že } L(M) \neq \emptyset\}$$

- Navrheme redukci (redukující HP na NEMP, viz směr 2, „NEMP je aspoň tak těžký jako HP“):

$$HP \leq NEMP$$

$$\sigma : \{0, 1, \#\}^* \rightarrow \{0, 1\}^*$$

- Funkce  $\sigma$  přiřadí řetězci  $x \in \{0, 1, \#\}^*$  řetězec  $\langle M_\sigma \rangle$ , kde  $M_\sigma$  je TS implementující  $\sigma$ , pracující následovně:

- $M_\sigma$  smaže svůj vstup (pozn. jazyk  $L(M_\sigma)$  tedy bude bud'  $\emptyset$  a nebo  $\Sigma^*$ ).
- $M_\sigma$  zapíše na svůj vstup řetěze  $x$ .
- $M_\sigma$  ověří, zda  $x = \{\langle M \rangle \# \langle w \rangle\}$  má validní strukturu pro TS  $M$  a jeho vstup  $w$ . Pokud nemá, tak odmítne.
- TS  $M_\sigma$  odsimuluje běh TS  $M$  na  $w$ . Pokud  $M$  zastaví, tak přijme, jinak cyklí.

- Funkce  $\sigma$  lze evidentně realizovat úplným TS  $M_\sigma$ . Skládá se, ze 4 komponent:

1. Smazání vstupu – konstantní operace, pro kterou  $M_\sigma$  pouze vypíše daný kód.
2. Zapsání  $x$  na vstup – snadná operace,  $M_\sigma$  vypíše kód, který provede přesun doprava a zápis  $a_i$  pro  $x = a_1 a_2 \dots a_n$  pro  $i = 1, 2, \dots, n$ .
3. Ověření správného strukturování vstupu – konstantní operace, pro kterou  $M_\sigma$  pouze vypíše daný kód.
4. Předání řízení univerzálnímu TS – konstantní operace, pro kterou  $M_\sigma$  pouze vypíše daný kód.

- Ukázání zachování členství v jazyce.

- Studujme jazyk  $M_\sigma$ :

- \*  $L(M_\sigma) = \emptyset \Leftrightarrow x$  nemá validní strukturu  $\langle M \rangle \# \langle w \rangle$  a nebo  $M$  na  $w$  nezastaví.
- \*  $L(M_\sigma) = \Sigma^* \Leftrightarrow x$  má validní strukturu  $\langle M \rangle \# \langle w \rangle$  a  $M$  na  $w$  zastaví.

- Zachování členství:

$$\begin{aligned} \forall x \in \{0, 1, \#\}^* : \sigma(x) = \langle M_\sigma \rangle \in NEMP &\Leftrightarrow \\ &\Leftrightarrow L(M_\sigma) = \Sigma^* \Leftrightarrow \\ &\Leftrightarrow x = \langle M \rangle \# \langle w \rangle \text{ a } M \text{ na } w \text{ zastaví} \Leftrightarrow \\ &\Leftrightarrow x \in HP \end{aligned}$$

**Příklad 2: dokažte, že problém prázdnosti jazyka TS není ani částečně rozhodnutelný**

- Dokážeme redukcí z problému nezastavení TS (coHP).
- Jazyk charakterizující problém nezastavení (obsahuje takové instance, pro které TS  $M$  na  $w$  nezastaví):

$$coHP = \{\langle M \rangle \# \langle w \rangle \mid \text{TS } M \text{ na vstupu } w \text{ nezastaví}\}$$

- Jazyk charakterizující problém prázdnosti (obsahuje takové instance, pro které jazyk TS  $M$  je prázdný):

$$EMP = \{\langle M \rangle \mid M \text{ je TS takový, že } L(M) = \emptyset\}$$

- Navrhne redukci (redukující coHP na EMP, viz směr 2, „EMP je aspoň tak těžký jako coHP“):

$$coHP \leq EMP$$

$$\sigma : \{0, 1, \#\}^* \rightarrow \{0, 1\}^*$$

- Funkce  $\sigma$  přiřadí řetězci  $x \in \{0, 1, \#\}^*$  řetězec  $\langle M_\sigma \rangle$ , kde  $M_\sigma$  je TS implementující  $\sigma$ , pracující následovně:

- $M_\sigma$  smaže svůj vstup (pozn. jazyk  $L(M_\sigma)$  tedy bude bud'  $\emptyset$  a nebo  $\Sigma^*$ ).
- $M_\sigma$  zapíše na svůj vstup řetězez  $x$ .
- $M_\sigma$  ověří, zda  $x = \{\langle M \rangle \# \langle w \rangle\}$  má validní strukturu pro TS  $M$  a jeho vstup  $w$ . Pokud nemá, tak přijme.
- TS  $M_\sigma$  odsimuluje běh TS  $M$  na  $w$ . Pokud  $M$  zastaví, tak přijme, jinak cyklí.

- Zbytek na stejný princip jako v příkladu 1.

**Příklad 3: dokažte, že problém náležitosti řetězce jazyku  $L(M)$ , kde  $M$  je TS, je nerozhodnutelný**

- Dokážeme redukcí z problému zastavení TS (HP).
- Jazyk charakterizující problém zastavení (obsahuje takové instance, pro které TS  $M$  na  $w$  zastaví):

$$HP = \{\langle M \rangle \# \langle w \rangle \mid \text{TS } M \text{ na vstupu } w \text{ zastaví}\}$$

- Jazyk charakterizující problém náležitosti řetězce jazyku  $L(M)$ , kde  $M$  je TS (obsahuje takové instance, pro které jazyk TS  $M$  na  $w$  zastaví a přijme):

$$MP = \{\langle M \rangle \# \langle w \rangle \mid M \text{ je TS takový, že } w \in L(M)\}$$

- Navrhne redukci (redukující HP na MP, viz směr 2, „MP je aspoň tak těžký jako HP“):

$$HP \leq MP$$

$$\sigma : \{0, 1, \#\}^* \rightarrow \{0, 1, \#\}^*$$

$$\sigma(\langle M \rangle \# \langle w \rangle) = \langle M' \rangle \# \langle w' \rangle$$

- Funkce  $\sigma$  přiřadí řetězci  $x \in \{0, 1, \#\}^*$  řetězec  $\langle M \rangle \# \langle w \rangle$ , kde  $M_\sigma$  je TS implementující  $\sigma$ , pracující následovně:

- $M_\sigma$  smaže svůj vstup (pozn. jazyk  $L(M_\sigma)$  tedy bude bud'  $\emptyset$  a nebo  $\Sigma^*$ ).
- $M_\sigma$  zapíše na svůj vstup řetězez  $x$ .
- $M_\sigma$  ověří, zda  $x = \{\langle M \rangle \# \langle w \rangle\}$  má validní strukturu pro TS  $M$  a jeho vstup  $w$ . Pokud nemá, tak [[todo: přijme?]].
- TS  $M_\sigma$  odsimuluje běh TS  $M$  na  $w$ . Pokud  $M$  na  $w$  zastaví přijme, tak  $M_\sigma$  přijme. Jinak,  $M_\sigma$  cyklí a nebo zamítá.

- Zbytek na stejný princip jako v příkladu 1.

Příklad 4: dokažte, že problém víceznačnosti bezkontextových gramatik, není ani částečně rozhodnutelný

[**note:** Příklad je na redukci z PCP, jedná se o těžší redukci, která se nebude zkoušet.]

## Kapitola 7

# TIN – Časová a paměťová složitost (třídy složitosti, úplnost, SAT problém).

### 7.1 Zdroje

- tin\_2021\_merged.pdf
- TIN\_todo.mp4

### 7.2 Úvod a kontext

[[todo]]

## Kapitola 8

# GAL – Hledání minimální kostry obyčejného grafu (pojmy, stromy a kostry, Kruskalův algoritmus, Primův algoritmus).

### 8.1 Zdroje

- gal-handouts.pdf
- GAL\_2020-10-22.mp4
- GAL\_2020-10-29.mp4

### 8.2 Úvod a kontext

**Orientovaný graf** Orientovaný graf je dvojice  $G = (V, E)$ , kde  $V$  je konečná množina uzlů a  $E \subseteq V \times V$  je množina hran.

**Neorientovaný graf** Neorientovaný graf je dvojice  $G = (V, E)$ , kde  $V$  je konečná množina uzlů a  $E \subseteq \binom{V}{2}$  je množina hran. (Hrana je tedy dvouprvková množina, avšak běžně se držíme stejného značení jako u orientovaných grafů a používáme dvojici.)

**Ohodnocený graf** Ohodnocený graf je takový graf, jehož každá hrana má přiřazenou nějakou hodnotu, typicky definovanou pomocí váhové funkce  $w : E \mapsto \mathbb{R}$ .

**Podgraf** Graf  $G' = (V', E')$  je podgraf grafu  $G = (V, E)$  jestliže  $V' \subseteq V$  a  $E' \subseteq E$ .

**Sled** Posloupnost uzlů  $\langle v_0, v_1, \dots, v_k \rangle$ , kde  $(v_{i-1}, v_i) \in E$  pro  $i = 1, \dots, k$  se nazývá sled délky  $k$  z  $v_0$  do  $v_k$ .

**Uzavřený sled** Sled  $\langle v_0, v_1, \dots, v_k \rangle$  se nazývá uzavřený, pokud existuje hrana  $(v_0, v_k)$ .

**Dosažitelnost** Pokud existuje sled  $s$  z uzlu  $u$  do uzlu  $v$ , říkáme, že  $v$  je dosažitelný z  $u$  sledem  $s$ , značeno  $u \xrightarrow{s} v$ .

**Tah** Tah je sled ve kterém se neopakují hrany.

**Cesta** Cesta je sled ve kterém se neopakují uzly.

**Souvislý graf** Neorientovaný graf se nazývá souvislý, pokud mezi libovolnými dvěma uzly existuje cesta.

**Kružnice** Uzavřená cesta se nazývá kružnice.

**Cyklus** Orientovaná kružnice se nazývá cyklus (první a poslední uzel je shodný).

**Prostý graf** Orientovaný graf bez cyklů se nazývá prostý.

**Acyklický graf** Graf je bez cyklů, resp. kružnic, se nazývá acyklický.

**Strom** Graf, který je souvislý a acyklický, se nazývá strom.

**Kostra** Strom, který tvoří podgraf souvislého grafu na množině všech jeho vrcholů, se nazývá kostra (*spanning tree*).

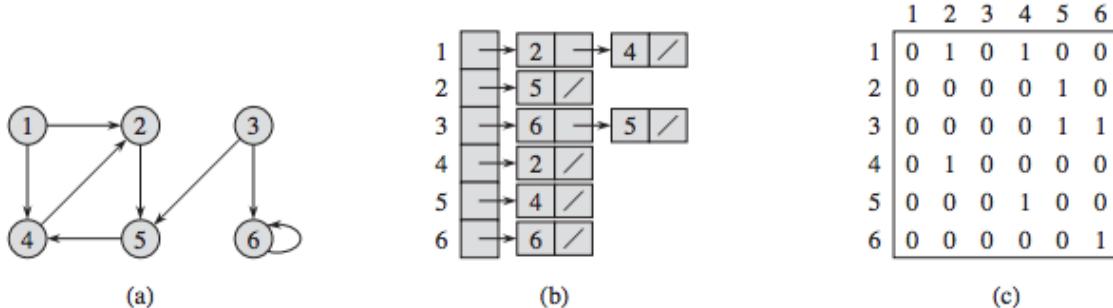
**Minimální kostra** Necht'  $G = (V, E)$  je souvislý neorientovaný graf s váhovou funkcí  $w : E \mapsto \mathbb{R}$ . Minimální kostra (*MST, minimum spanning tree*) je strom  $G' = (V, E')$ , kde  $E' \subseteq E$  a

$$w(E') = \sum_{(u,v) \in T} w(u,v)$$

je minimální ze všech možných alternativních koster.

**Seznam sousedů** Seznam sousedů (*Adj, adjacency list*) je reprezentace grafu v paměti. Jde o preferovanou variantu pro řídké grafy – kde  $m \ll n^2$ . Pro každý uzel máme definovaný seznam jeho sousedů.

**Matice sousednosti** Matice sousednosti (*adjacency matrix*) je reprezentace grafu v paměti. Jde o preferovanou variantu pro husté grafy – kde  $m$  je skoro  $n^2$ .



Obrázek 8.1: Příklad reprezentace grafu pomocí seznamu sousedů a matice sousednosti.

### 8.3 Generický algoritmus

Hledání minimální kostry je problém, který lze řešit algoritmy, které spadají do kategorie tzv. hladových (*greedy*) deterministických algoritmů. Spočívají v tom, že průběžně odhadují kostru přidáváním dalších hran a nikdy se nemusejí vracet (neprovádí se *backtracking*). Generický algoritmus tvoří jakousi základní kostru pro další, už konkrétní, algoritmy.

**Řez** Necht'  $G = (V, E)$  je graf. Řez grafu  $G$  je dvojice  $(S, V - S)$ , kde  $\emptyset \subseteq S \subseteq V$ .

**Křížení** Hrana  $(u, v) \in E$  kříží řez  $(S, V - S)$ , pokud jeden její konec je v  $S$  a druhý v  $V - S$ .

**Respektování** Necht'  $A \subseteq E$  je množina hran. Řez  $(S, V - S)$  respektuje množinu hran  $A$ , pokud žádná hrana v  $A$  nekříží řez  $(S, V - S)$ .

**Lehkost** Necht'  $(S, V - S)$  je řez a  $B$  je množina hran, která ho kříží. Hrana z množiny  $B$  s nejmenší hodnotou se nazývá lehká.

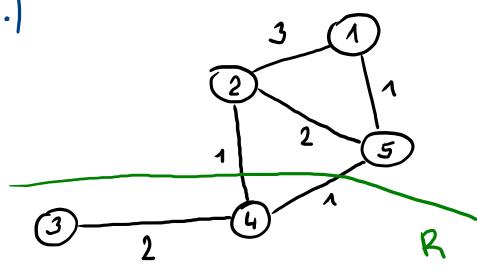
**Bezpečnost** Necht'  $G = (V, E)$  je souvislý neorientovaný graf s reálnou váhovou funkcí  $w$ . Necht'  $A \subseteq E$  je součástí nějaké minimální kostry  $G$ . Necht'  $(S, V - S)$  je řez, který respektuje  $A$ . Necht'  $(u, v)$  je lehká hrana křížící  $(S, V - S)$ . Pak hrana  $(u, v)$  je bezpečná pro  $A$ .

```
1 def generic_mst(G):
2     # G je graf
3     A = {}# A je mnozina hran rozpracovane minimalni kostry
4     while netvori_kostru(A, G):
5         for hrana in G.E:
6             if je_bezpecna(A, hrana):
7                 A += {hrana}
8     return A
```

Výpis 8.1: Generický algoritmus. Před každou iterací algoritmu je množina  $A$  podmnožinou nějaké minimální kostry. Hrana  $(u, v) \in E$  je bezpečná pro  $A$ , pokud  $A \cup \{(u, v)\}$  je podmnožinou nějaké minimální kostry.

Př.  $G = (V, E)$   $\psi: E \rightarrow \mathbb{R}$  (definované obrazem)

1.)



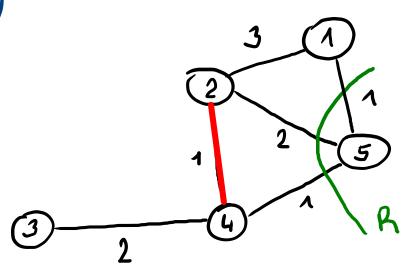
$$A = \{\}$$

$$R = (\{\{1, 2, 5\}, \{3, 4\}\})$$

$$LH = \{(2, 4), (4, 5)\}$$

$$A \leftarrow A \cup \{(2, 4)\}$$

2.)



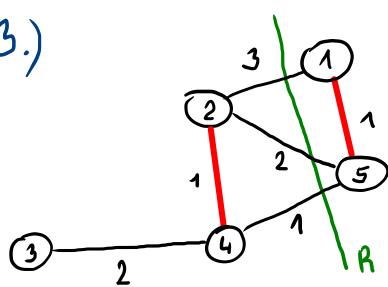
$$A = \{(2, 4)\}$$

$$R = (\{\{1, 2, 3, 4\}, \{5\}\})$$

$$LH = \{(1, 3), (4, 5)\}$$

$$A \leftarrow A \cup \{(1, 3)\}$$

3.)



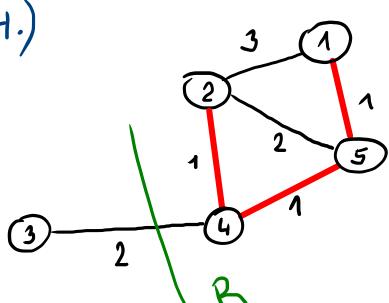
$$A = \{(1, 3), (2, 4)\}$$

$$R = (\{\{1, 3\}, \{2, 3, 4\}\})$$

$$LH = \{(4, 5)\}$$

$$A \leftarrow A \cup \{(4, 5)\}$$

4.)



$$A = \{(1, 3), (2, 4), (4, 5)\}$$

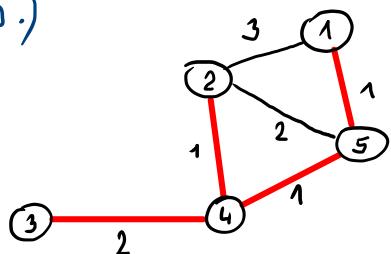
$$R = (\{\{1, 2, 4, 5\}, \{3\}\})$$

$$LH = \{(3, 4)\}$$

$$A \leftarrow A \cup \{(3, 4)\}$$

Obrázek 8.3: Příklad, část 2.

5.)



$$A = \{(3,4), (2,4), (5,4), (1,5)\}$$

*A je minimální kostra*

*(minimál už nemůže učítat  
věz, když by vyspetoval A)*

Obrázek 8.4: Příklad, část 3.

## 8.4 Kruskalův algoritmus

Kruskalův a Primův algoritmus se liší v tom, jakým způsobem vybírají bezpečnou hranu. Kruskalův algoritmus nahlíží na  $A$  jako na les a hledá hranu s nejmenším ohodnocením, která spojuje stromy v lese. Na konci je  $A$  jeden strom.

```

1 def kruskal_mst(G):
2     # G je graf
3
4     # inicializace, kazdy uzel je ve sve mnozine
5     A = {}# A je mnozina hran rozpracovane minimalni kostry
6     for v in G.V:
7         make_set(v)
8
9     # seradit vzestupne podle w
10    E = sort(G.E, G.w)
11
12    for (u, v) in E:
13        if find_set(u) != find_set(v):
14            A += {(u, v)}
15            union(u, v)
16
17    return A

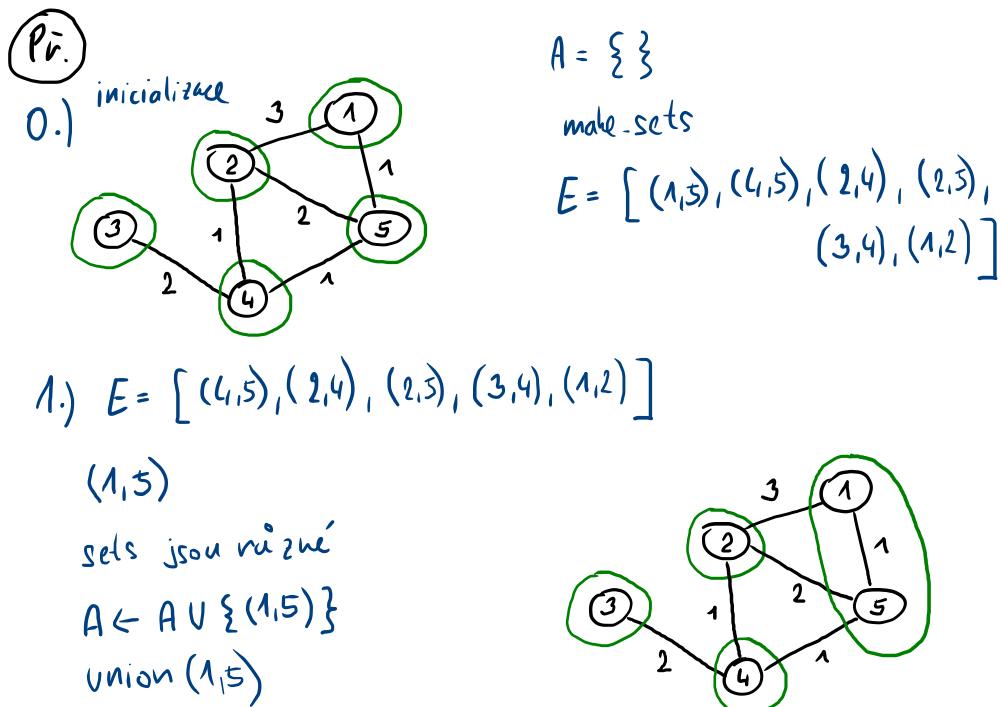
```

Výpis 8.2: Kruskalův algoritmus. Funkce `make_set(v)` vytvoří množinu obsahující  $v$ , `find_set(v)` vrátí reprezentanta množiny ve které se nachází  $v$ , `union(u, v)` sjednotí dvě množiny obsahující  $u$  a  $v$ .

#### 8.4.1 Složitost

- Řádek 5 –  $O(1)$
- Řádek 6-7 –  $n$ -krát složitost *make\_set* ( $n$  je počet uzlů).
- Řádek 10 –  $O(m \cdot \log(m))$  ( $m$  je počet hran).
- Řádky 12-15 – Závisí na implementaci *find\_set* a *union*.
  - Při implementaci seznamem s heuristickou celkem:  $O(m + n \cdot \log(n))$ .
  - Při stromové implementaci s váhami a zkratkami celkem:  $O((m + n) \cdot \alpha(n))$ .  
Kde  $\alpha$  je velmi pomalu rostoucí funkce ( $\alpha \leq 4$ ).
- Pro souvislý graf platí  $m > n$ . Proto množinové operace stojí  $O(m \cdot \alpha(n))$ . Jelikož  $\alpha(n) = O(\log(n)) = O(\log(m))$ , tak celková složitost je  $O(m \cdot \log(m))$ .
- Dále platí  $m < n^2$ , pak  $\log(m) = O(\log(n))$ , proto celkem:  $O(m \cdot \log(n))$ .

#### 8.4.2 Příklad



Obrázek 8.5: Příklad, část 1.

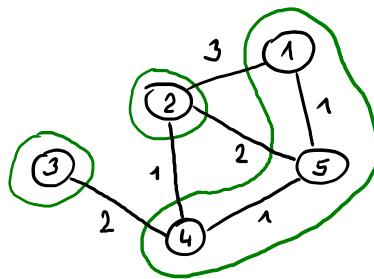
$$2.) E = [(2,4), (2,5), (3,4), (1,2)]$$

(4,5)

sets jsou různé

$$A \leftarrow A \cup \{(4,5)\}$$

union (4,5)



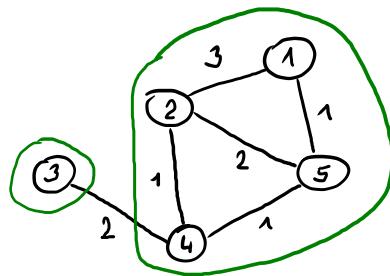
$$3.) E = [(2,3), (3,4), (1,2)]$$

(2,4)

sets jsou různé

$$A \leftarrow A \cup \{(2,4)\}$$

union (2,4)

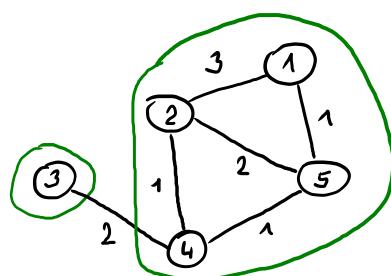


Obrázek 8.6: Příklad, část 2.

$$4.) E = [(3,4), (1,2)]$$

(2,3)

sets nejsou různé



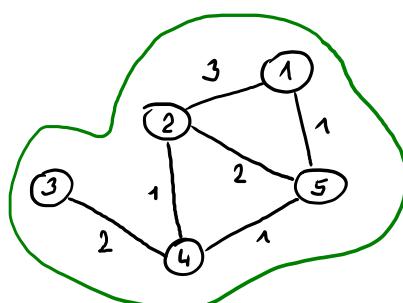
$$4.) E = [(1,2)]$$

(3,4)

sets jsou různé

$$A \leftarrow A \cup \{(3,4)\}$$

union (3,4)



$$5.) E = [ ] , (1,2) , \text{ sets nejsou různé}$$

Obrázek 8.7: Příklad, část 3.

## 8.5 Primův-Jarníkův algoritmus

Primův algoritmus buduje tzv.  $A$  strom. Má zadaný určitý uzel, ze kterého hledá nejbližší další uzel, který by připojil. A pak další a další.

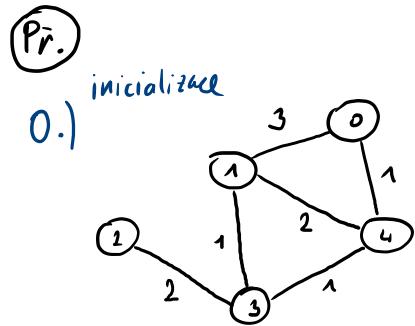
```
1 def prim_mst(G, r):
2     # G je graf
3     # r je výchozí uzel
4
5     for u in G.V:
6         key[u] = INF # pole cen prechodu, kolik stojí prechod do vrcholu na indexu
7         pi[u] = NULL # pole predchudcu, kdo je predchudce vrcholu na indexu
8
9     key[r] = 0
10    Q = Queue(G.V) # prioritní fronta uzlu
11
12    while not Q.empty():
13        u = Q.extract_min(key) # vrati prvek z Q s nejmensi hodnotou v key
14
15        # pro vsechny sousedy uzlu u (Adj je seznam sousedu)
16        for v in Adj[u]:
17            # pokud je levnejsi cesta a jeste to není prozkoumaný uzel
18            if v in Q and w(u, v) < key(v):
19                pi[v] = u
20                key[v] = w(u, v)
21                Q.decrease_key(key) # aktualizace prioritni fronty
22
23    return pi
```

Výpis 8.3: Primův algoritmus.

### 8.5.1 Složitost

- Řádky 5-10 –  $O(n)$  za použití binární haldy ( $n$  je počet uzlů).
- Řádky 12-13 – While cyklus se provede  $n$ -krát a protože  $extract\_min$  stojí  $O(\log(n))$ , tak je celková složitost  $O(n \cdot \log(n))$ .
- Řádek 16 – For cyklus se provede  $O(m)$  krát, protože délka všech seznamů sousedů je dohromady  $2m$  ( $m$  je počet hran).
- Řádek 18-20 –  $O(1)$ .
- Řádek 21 –  $O(\log(n))$ .
- Jelikož  $m > n$ , tak celkem  $O(n \cdot \log(n) + m \cdot \log(n)) = O(m \cdot \log(n))$ .

### 8.5.2 Příklad



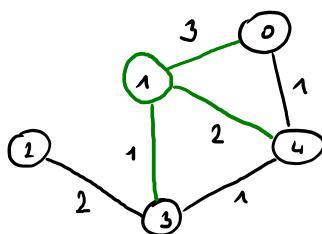
$$n = 1$$

$$\text{key} = [\infty, 0, \infty, \infty, \infty]$$

$$\pi = [\text{NULL}, \text{NULL}, \text{NULL}, \text{NULL}, \text{NULL}]$$

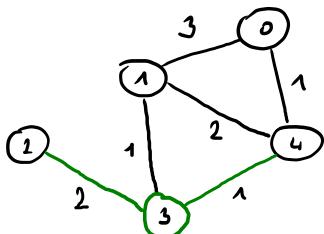
$$Q = [0, 1, 2, 3, 4]$$

1.)  $u = 1$   
 $Q = [0, 1, 3, 4]$   
 $v \in \{0, 3, 4\}$   
 $\text{key} = [3, 0, \infty, 1, 2]$   
 $\pi = [1, \text{NULL}, \text{NULL}, 1, 1]$

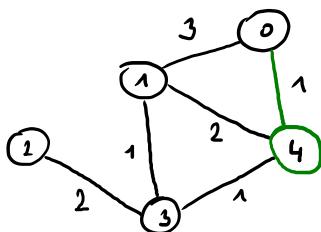


Obrázek 8.8: Příklad, část 1.

2.)  $u = 3$   
 $Q = [0, 1, 4]$   
 $v \in \{1, 2, 4\}$   
 $\text{key} = [3, 0, 2, 1, 1]$   
 $\pi = [1, \text{NULL}, 3, 1, 3]$



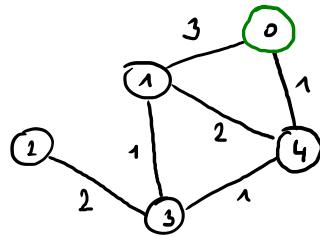
3.)  $u = 4$   
 $Q = [0, 1]$   
 $v \in \{0, 1, 3\}$   
 $\text{key} = [1, 0, 2, 1, 1]$   
 $\pi = [4, \text{NULL}, 3, 1, 3]$



Obrázek 8.9: Příklad, část 2.

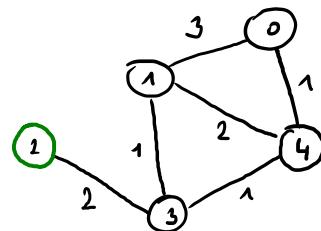
4.)  $u = 0$   
 $Q = [1]$   
 $v \in \{1, 4\}$

$key = [1, 0, 2, 1, 1]$   
 $\pi = [4, \text{NULL}, 3, 1, 3]$



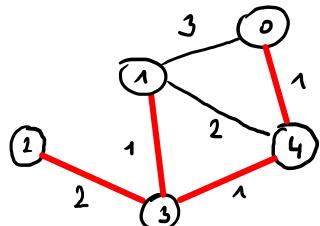
5.)  $u = 2$   
 $Q = []$   
 $v \in \{3\}$

$key = [1, 0, 2, 1, 1]$   
 $\pi = [4, \text{NULL}, 3, 1, 3]$



Obrázek 8.10: Příklad, část 3.

6.)  $key = [1, 0, 2, 1, 1]$   
 $\pi = [4, \text{NULL}, 3, 1, 3]$



Obrázek 8.11: Příklad, část 4.

# Kapitola 9

**GAL – Hledání nejkratších cest ze zdrojového uzlu do všech ostatních uzel grafu (Bellman-Fordův algoritmus, Dijkstrův algoritmus).**

## 9.1 Zdroje

- gal-handouts.pdf
- GAL\_2020-11-05.mp4

## 9.2 Úvod a kontext

**Cena cesty** Necht'  $G = (V, E)$  je ohodnocený graf s váhovou funkcí  $w : E \mapsto \mathbb{R}$ . Cena cesty  $p = \langle v_o, v_1, \dots, v_k \rangle$  je suma

$$w(p) = \sum_{i=0}^k w(v_i, v_{i+1})$$

**Cena nejkratší cesty** Cena nejkratší cesty z  $u$  do  $v$  je

$$\delta(u, v) = \begin{cases} \min(\{w(p) : u \xrightarrow{p} v\}) \\ \infty \text{ pokud cesta neexistuje} \end{cases}$$

**Nejkratší cesta** Nejkratší cesta z  $u$  do  $v$  je pak libovolná cesta  $p$  taková, že  $w(p) = \delta(u, v)$ .

**Cena cesty se záporným cyklem** Pokud na cestě z  $u$  do  $v$  existuje záporný cyklus (cyklus jehož celková cena je záporná), pak  $\delta(u, v) = -\infty$ .

**Záporné ohodnocení hran** Pokud na cestě z  $u$  do  $v$  neexistuje záporný cyklus, tak algoritmy pracují dobře i se záporným ohodnocením hran.

**Reprezentace cesty** Cestu reprezentujeme pomocí pole předchůdců  $\pi$ .

**Hledání nejkratších cest ze všech uzlů do jednoho** Tento problém lze řešit stejnými algoritmy. Graf se transponuje (převrácení orientace hran), provede se algoritmus pro problém „hledání nejkratších cest ze jednoho uzlu do všech ostatních uzlů“ a poté se transponuje zpět.

**Reprezentace nejkratší cesty** Nejkratší cestu grafu  $G = (V, E)$  reprezentujeme pomocí pole předchůdců  $\pi$ , kde  $\pi[v]$  označuje předchůdce uzlu  $v \in V$  na nejkratší cestě. Podgraf předchůdců pak je  $G_\pi = (V_\pi, E_\pi)$ ,  $V_\pi = \{v \in V : \pi[v] \neq \text{NULL}\} \cup \{s\}$ ,  $E_\pi = \{(\pi[v], v) \in E : v \in V_\pi - \{s\}\}$ . V okamžiku dokončení algoritmu výpočtu nejkratších cest je  $G_\pi$  strom nejkratších cest. Tj. kořenový strom obsahující nejkratší cesty ze zdroje  $s$  do všech ostatních uzlů.

### 9.3 Pomocné funkce

Představené algoritmy pracují z důvodu efektivity se sledy a nikoliv s cestami (bylo by nutné stále kontrolovat, zda nebyla porušena podmínka cesty), ačkoliv je problém nazývá hledání nejkratší cesty.

```

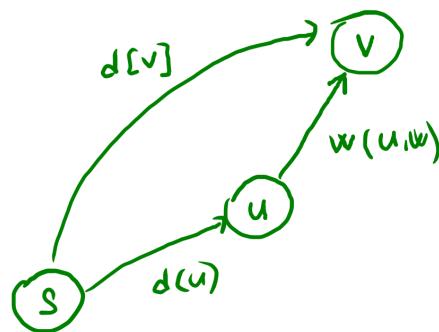
1 def initialize_single_source(G, s):
2     # G je graf
3     # s je výchozí uzel
4     for v in G.V:
5         d[v] = INF # d je pole vzdalenosti
6         pi[v] = NULL # pi je pole předchůdce
7         d[s] = 0
    
```

Výpis 9.1: Pomocná inicializační funkce. Složitost je  $\Theta(n)$ , kde  $n$  je počet uzlů.

```

1 def relax(u, v, w):
2     # u a v jsou uzly grafu
3     # w je vahova funkce
4     if d[v] > d[u] + w(u, v):
5         d[v] = d[u] + w(u, v)
6         pi[v] = u
    
```

Výpis 9.2: Pomocná funkce *relax*. Složitost je  $O(1)$ .



Obrázek 9.1: Ukázka činnosti funkce *relax*.

## 9.4 Bellman-Fordův algoritmus

Slouží pro řešení v obecných grafech, mohou obsahovat cykly a záporné hrany. Záporné cykly je však nutné detektovat a vrátit specifickou hodnotu. V podstatě se jedná o *brute force* algoritmus, provede se relaxace  $n - 1$ -krát pro každou hranu.

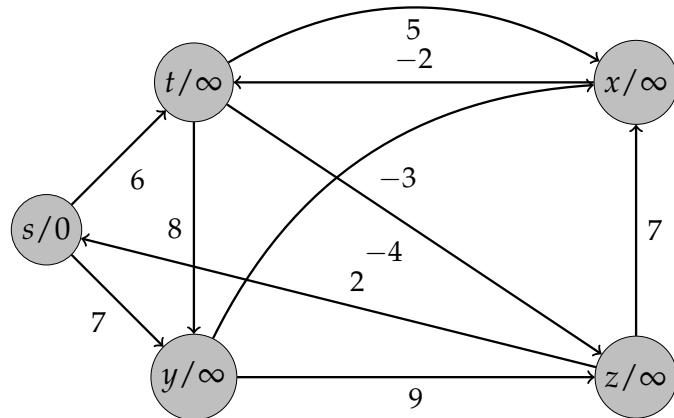
```
1 def bellman_ford(G, s, w):
2     # G je graf
3     # s je výchozí uzel
4     # w je vahova funkce
5
6     # faze inicializace
7     initialize_single_source(G, s)
8     n = len(G.V) # počet uzlu
9
10    # faze relaxace: provedení  $(n-1) * m$  relaxací ( $m$  je počet hran)
11    for _ in range(0, n-1):
12        for u, v in G.E:
13            relax(u, v, w)
14
15    # faze detekce záporného cyklu
16    for u, v in G.E:
17        if d[u] > d[v] + w(u, v):
18            return NULL
19
20    return pi
```

Výpis 9.3: Algoritmus Bellman-Ford. Proč  $n - 1$  iterací? Protože mezi libovolnými dvěma uzly v grafu, existuje cesta o maximálním počtu hran  $n - 1$ .

### 9.4.1 Složitost

- Řádek 7, 8 –  $\Theta(1)$ .
- Řádky 11, 12, 13 –  $(n - 1) \cdot \Theta(m) = \Theta(n \cdot m)$ , kde  $n$  je počet uzlů a  $m$  je počet hran grafu.
- Řádek 16, 17, 18 –  $\Theta(m)$ .
- Celkem  $\Theta(n \cdot m)$ .

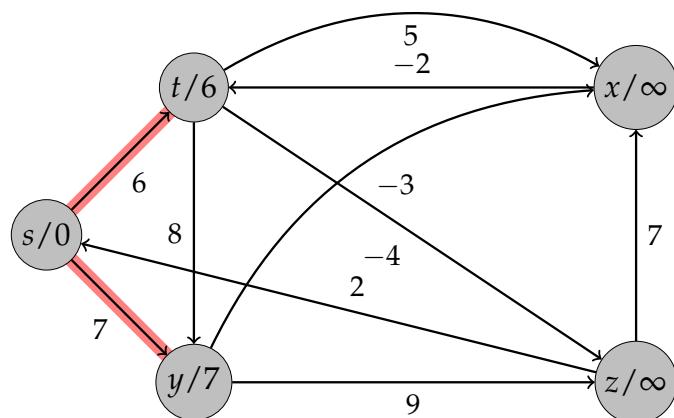
#### 9.4.2 Příklad



Obrázek: Práce algoritmu Bellman-Ford.

- ▶ Pokud  $(u, v) \in E$  je označená, pak  $\pi[v] = u$
- ▶ Hrany se relaxují v tomto pořadí:  
 $(t, x), (t, y), (t, z), (x, t), (y, x), (y, z), (z, x), (z, s), (s, t), (s, y)$ .

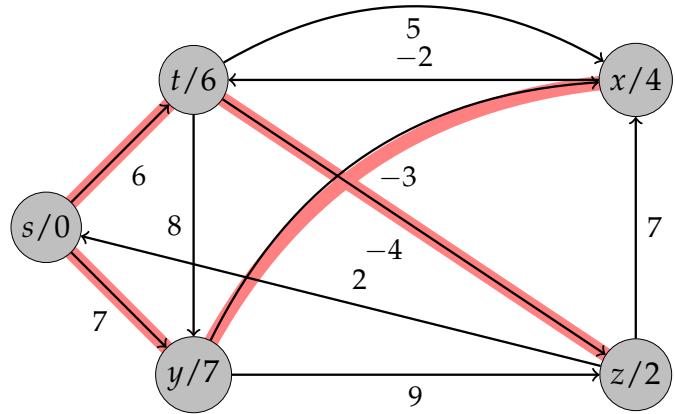
Obrázek 9.2: Příklad, část 1.



Obrázek: Práce algoritmu Bellman-Ford.

- ▶ Pokud  $(u, v) \in E$  je označená, pak  $\pi[v] = u$
- ▶ Hrany se relaxují v tomto pořadí:  
 $(t, x), (t, y), (t, z), (x, t), (y, x), (y, z), (z, x), (z, s), (s, t), (s, y)$ .

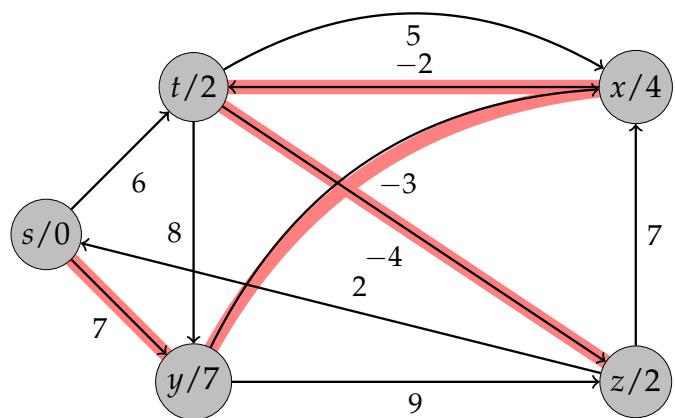
Obrázek 9.3: Příklad, část 2.



Obrázek: Práce algoritmu Bellman-Ford.

- ▶ Pokud  $(u, v) \in E$  je označená, pak  $\pi[v] = u$
- ▶ Hrany se relaxují v tomto pořadí:  
 $(t, x), (t, y), (t, z), (x, t), (y, x), (y, z), (z, x), (z, s), (s, t), (s, y)$ .

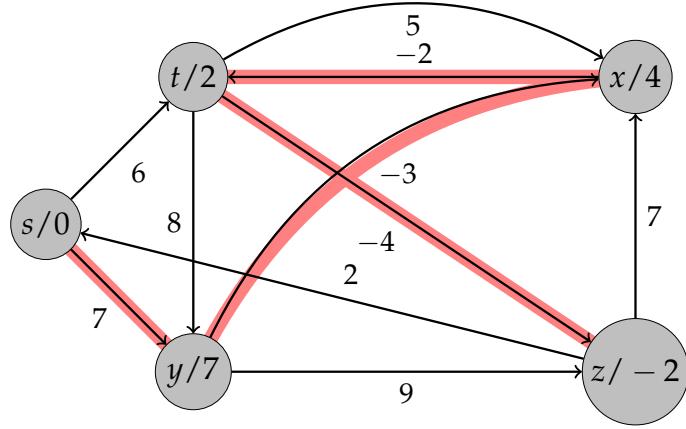
Obrázek 9.4: Příklad, část 3.



Obrázek: Práce algoritmu Bellman-Ford.

- ▶ Pokud  $(u, v) \in E$  je označená, pak  $\pi[v] = u$
- ▶ Hrany se relaxují v tomto pořadí:  
 $(t, x), (t, y), (t, z), (x, t), (y, x), (y, z), (z, x), (z, s), (s, t), (s, y)$ .

Obrázek 9.5: Příklad, část 4.



Obrázek: Práce algoritmu Bellman-Ford.

- ▶ Pokud  $(u, v) \in E$  je označená, pak  $\pi[v] = u$
- ▶ Hrany se relaxují v tomto pořadí:  
 $(t, x), (t, y), (t, z), (x, t), (y, x), (y, z), (z, x), (z, s), (s, t), (s, y)$ .

Obrázek 9.6: Příklad, část 5.

## 9.5 Dijkstrův algoritmus

Slouží pro řešení v acyklických grafech bez záporných hran. Pro takto omezený problém existují rychlejší algoritmy než pro problém v obecných grafech.

```

1 def dijkstra(G, s, w):
2     # G je graf
3     # s je výchozí uzel
4     # w je vahová funkce
5
6     # fáze inicializace
7     initialize_single_source(G, s)
8     Q = Queue(G.V) # prioritní fronta uzlu
9     S = {}# množina uzlu, která už byla prozkoumána
10
11    # fáze relaxace
12    while not Q.empty():
13        u = Q.extract_min(d) # vrátí prvek z Q s nejménší hodnotou v d
14        S += {u}
15        # pro všechny sousedy uzlu u (Adj je seznam sousedů)
16        for v in Adj[u]:
17            relax(u, v, w)
18
19        Q.decrease_key(d) # aktualizace prioritní fronty
20
21    return d, pi

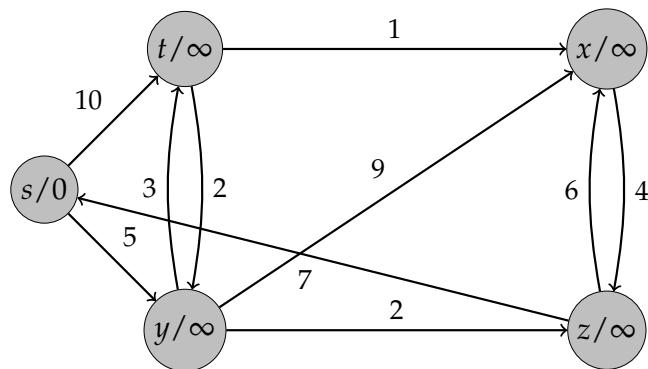
```

Výpis 9.4: Algoritmus Dijkstra.

### 9.5.1 Složitost

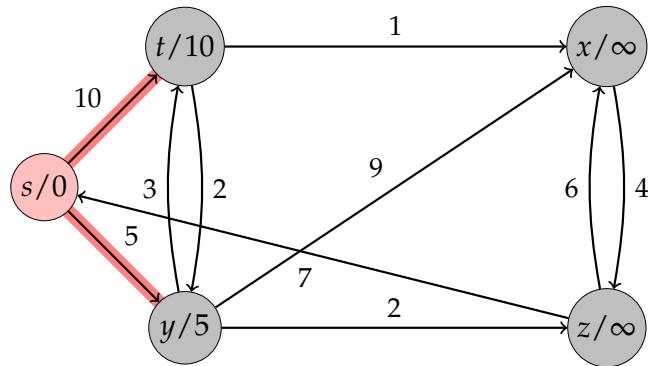
- Předpokládejme implementaci prioritní fronty pomocí pole.
- Řádek 8, 18 –  $O(1)$ .
- Řádek 11 – While cyklus se provede  $n$ -krát, kde  $n$  je počet uzlů.
- Řádek 12 –  $O(n)$ , najítí minima v poli uzlů. Celkově (s cyklem)  $O(n^2)$ .
- Řádek 16 –  $O(m)$ , pro všechny hrany. Celkově (s cyklem)  $O(m \cdot n)$ .
- Celkem  $O(n^2 + m) = O(n^2)$ .
- Pro řídké grafy lze využít implementaci fronty pomocí binární haldy a získat tak  $O(m \cdot \log(n))$ .
- Při implementaci fronty pomocí Fibonacciho haldy dostaneme časovou složitost  $O(n \cdot \log(n) + m)$ .

### 9.5.2 Příklad



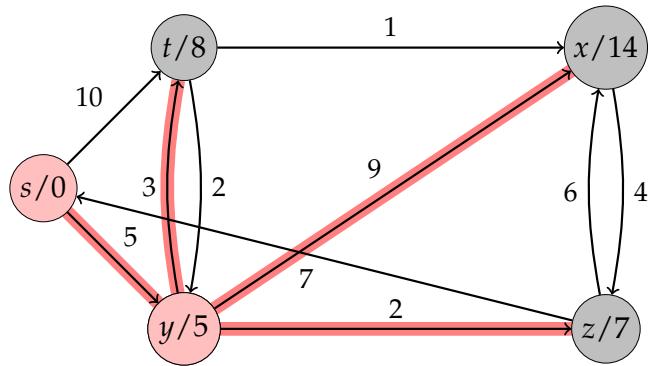
Obrázek: Práce Dijkstrova algoritmu. Označené uzly značí uzly z množiny  $S$ .

Obrázek 9.7: Příklad, část 1.



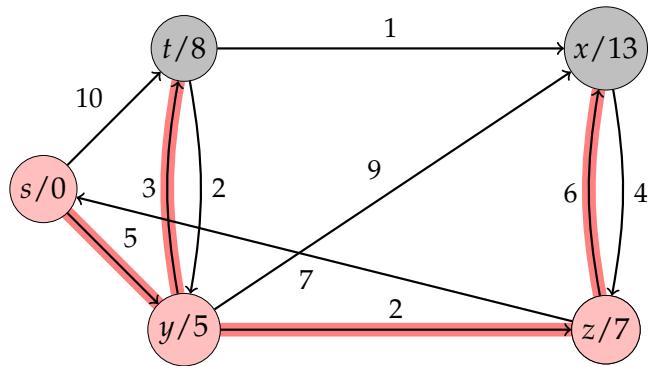
Obrázek: Práce Dijkstrova algoritmu. Označené uzly značí uzly z množiny  $S$ .

Obrázek 9.8: Příklad, část 2.



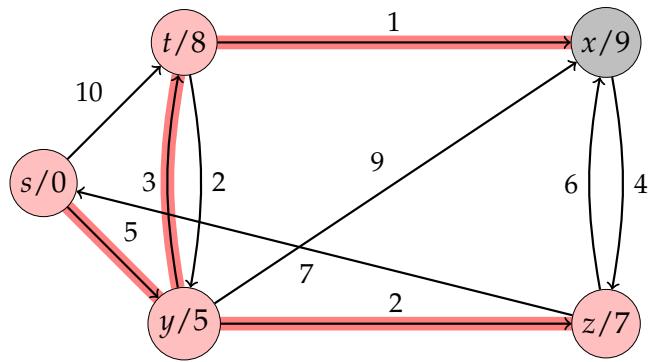
Obrázek: Práce Dijkstrova algoritmu. Označené uzly značí uzly z množiny  $S$ .

Obrázek 9.9: Příklad, část 3.



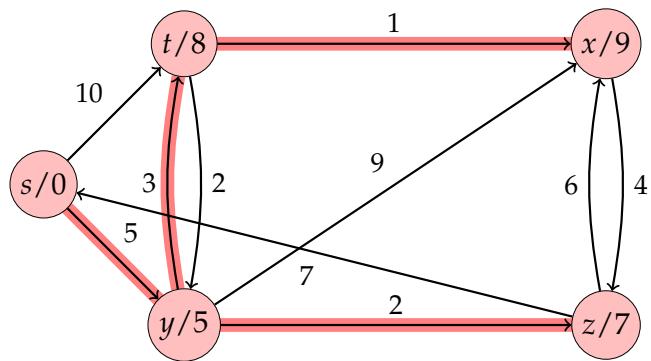
Obrázek: Práce Dijkstrova algoritmu. Označené uzly značí uzly z množiny  $S$ .

Obrázek 9.10: Příklad, část 4.



Obrázek: Práce Dijkstrova algoritmu. Označené uzly značí uzly z množiny  $S$ .

Obrázek 9.11: Příklad, část 5.



Obrázek: Práce Dijkstrova algoritmu. Označené uzly značí uzly z množiny  $S$ .

Obrázek 9.12: Příklad, část 6.

# Kapitola 10

## PDI – Klasifikace algoritmů volby koordinátora, algoritmus Bully a jeho složitost.

### 10.1 Zdroje

- 07\_Synchronization.pdf
- PDI\_2020-11-02.mp4

### 10.2 Úvod a kontext

- Mějme množinu procesů v rámci distribuovaného systému. Řešíme problém nalezení shody na nějaké věci (synchronizační problém). Problém můžeme rozdělit na dvě situace:
  - **Problém volby koordinátora** – Výběr jednoho z procesů, který bude vedoucím procesem (koordinátor). Tento proces pak může vykonat určitou činnost nebo může sloužit ostatním procesům k realizaci význačné role v systému.
  - **Problém vzájemného vyloučení** – Předpokládejme, že konkrétní zdroj může v daném okamžiku používat pouze jeden proces. Tento problém se běžně vyskytuje ve víceprocesorových systémech, ale také v distribuovaných systémech.
- Synchronizační problémy lze v rámci operačních systémů nebo multiprocesorových systémů řešit pomocí provádění atomických operací, sdílené paměti apod. (je pro ně podpora v rámci operačního systému nebo hardwaru). V distribuovaných systémech nic takového není z principu možné a proto se synchronizační problémy řeší pomocí zasílání zprav, resp. algoritmicky.

### 10.3 Problém volby koordinátora

- Předpokládáme:
  - Každý proces má unikátní ID.
  - Procesy neznají stav (běžící, neběžící) dalších procesů.
  - Každý proces zná ID dalších procesů (záleží na topologii).

- Cíl:
  - Dosáhnutí shody mezi všemi procesy na procesu, který je koordinátor.
  - Kritérium výběru koordinátora může být různé. Např. na základě proces ID (proces s největším ID se stane koordinátorem).

## 10.4 Bully algoritmus

Pro topologii každý s každým – každý proces může komunikovat s každým dalším procesem. Používá tři druhy zpráv: ELECTION, OK, COORDINATOR.

### 10.4.1 Postup

- Proces P, který má podezření, že chybí koordinátor, může zahájit volby.
  1. Proces P odešle zprávu ELECTION všem procesům s větším ID.
  2. Pokud nikdo neodpoví, P vyhrává volby a stává se koordinátorem.
  3. Pokud některý z procesů s větším ID odpoví (zpráva OK), tak přebírá řízení a práce P je ukončena.
  4. Pokud P obdrží zprávu ELECTION od procesů s menším ID, pošle jim odpověď OK na zablokování procesů.
- Nakonec zůstane pouze P (nový koordinátor), který o tom informuje ostatní zasláním zprávy COORDINATOR.
- Pokud se proces probudí nebo je restartován, první akcí je vyvolání voleb.

### 10.4.2 Složitost

Složitost z hlediska počtu zpráv.

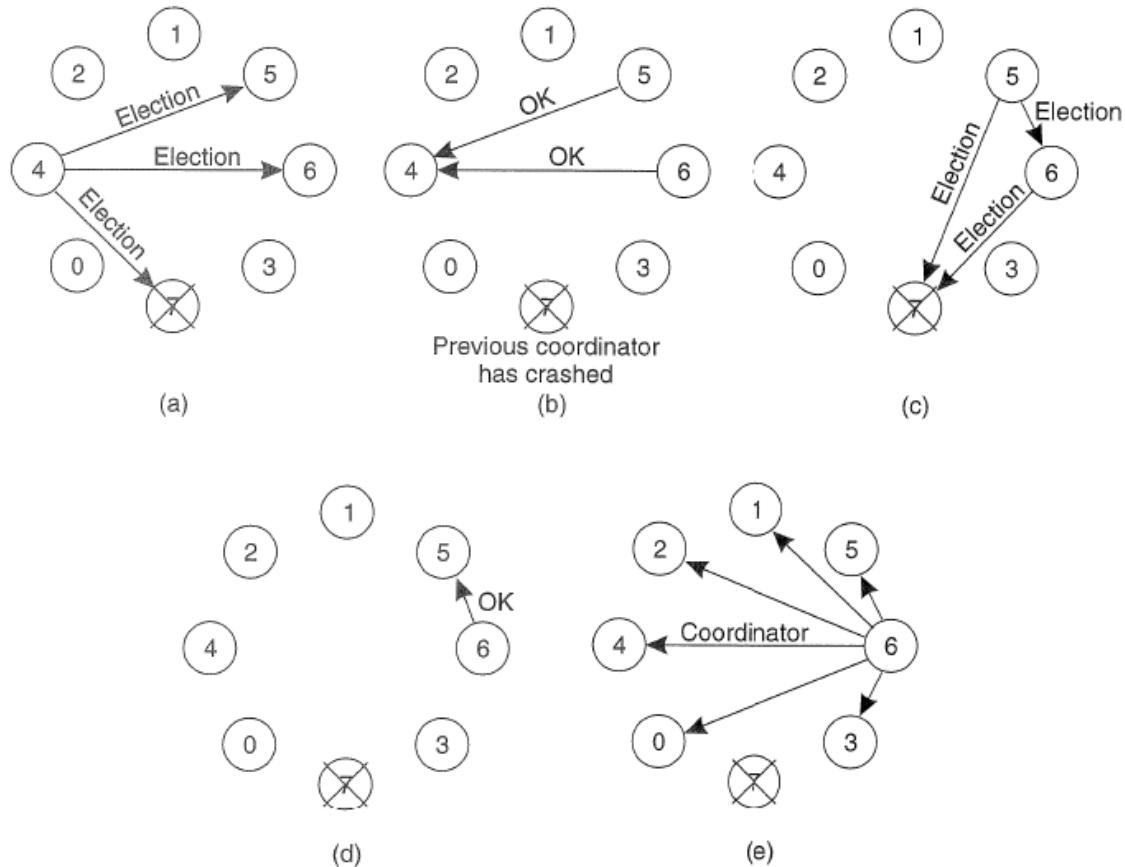
Nejhorský případ (iniciátor s nejmenším ID):

- $(n - 1)$  iterací
- $2(n - 1)$  zpráv ELECTION a OK pro každou iteraci
- $(n - 1)$  zpráv COORDINATOR
- Celkem:  $(n - 1) \times 2(n - 1) + (n - 1) \approx n^2$

Nejlepší případ (iniciátor s největším ID):

- $(n - 1)$  zpráv COORDINATOR
- Celkem:  $(n - 1)$

### 10.4.3 Příklad



**Figure 5-11.** The bully election algorithm. (a) Process 4 holds an election. (b) Processes 5 and 6 respond, telling 4 to stop. (c) Now 5 and 6 each hold an election. (d) Process 6 tells 5 to stop. (e) Process 6 wins and tells everyone.

Obrázek 10.1: Příklad činnosti Bully algoritmu.

## 10.5 Ring Algoritmus

Pro kruhovou topologii – procesy jsou uspořádané do kruhu podle svého proces ID. Každý proces musí vědět nejenom o svém následovníkovi, ale také o jeho následníkovi, který funguje jako „záloha“, v případě že by se přímý následník stal nedostupný. Používá dva druhy zpráv: ELECTION, COORDINATOR.

### 10.5.1 Postup

- Proces P, který má podezření, že chybí koordinátor, může zahájit volby.
  1. Zašle zprávu ELECTION obsahující jeho ID dalšímu procesu (pokud další proces nereaguje, proces P zašle stejnou zprávu dalšímu v kruhu).
  2. Každý člen topologie přijme zprávu ELECTION, přidá do ní své ID a přepošle zprávu dalšímu procesu.

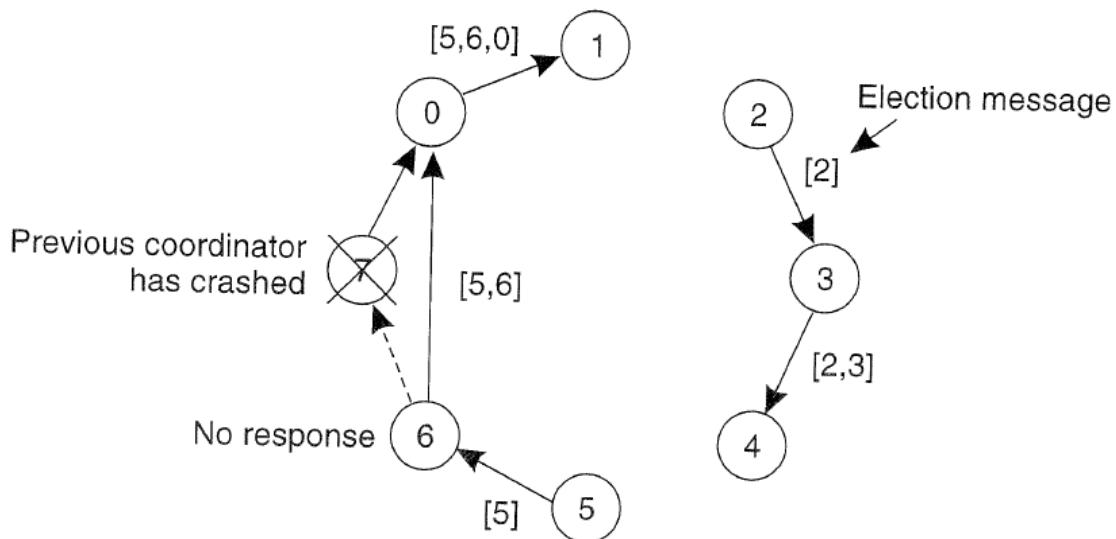
- Když se zpráva vrátí k procesu P, je zpráva převedena na zprávu COORDINATOR a poslána následujícímu procesu v topologii, aby bylo možné nahlásit:
  1. Novým koordinátorem se stává proces s nejvyšším ID.
  2. Členové sítě jsou stále aktivní.
- Po síti může obíhat více zpráv zároveň.

### 10.5.2 Složitost

Složitost z hlediska počtu zpráv.

Vždy  $2n \approx n$  zpráv. Jedno kolečko „oběhne“ zpráva ELECTION a druhé zpráva COORDINATOR.

### 10.5.3 Příklad



**Figure 5-12.** Election algorithm using a ring.

Obrázek 10.2: Příklad činnosti Ring algoritmu.

## 10.6 Algoritmus pro obecnou topologii

Předpokládáme, že nemáme ani kruhovou topologii ani spojení každý s každým. Např.: peer-to-peer sítě, sensorové sítě, ...

### 10.6.1 Postup

- V první iteraci se broadcastem posílá zpráva ELECTION.
- Každý uzel si uloží od kterého souseda dostal zprávu ELECTION jako první. Tím vzníká kostra grafu (*spanning tree*).

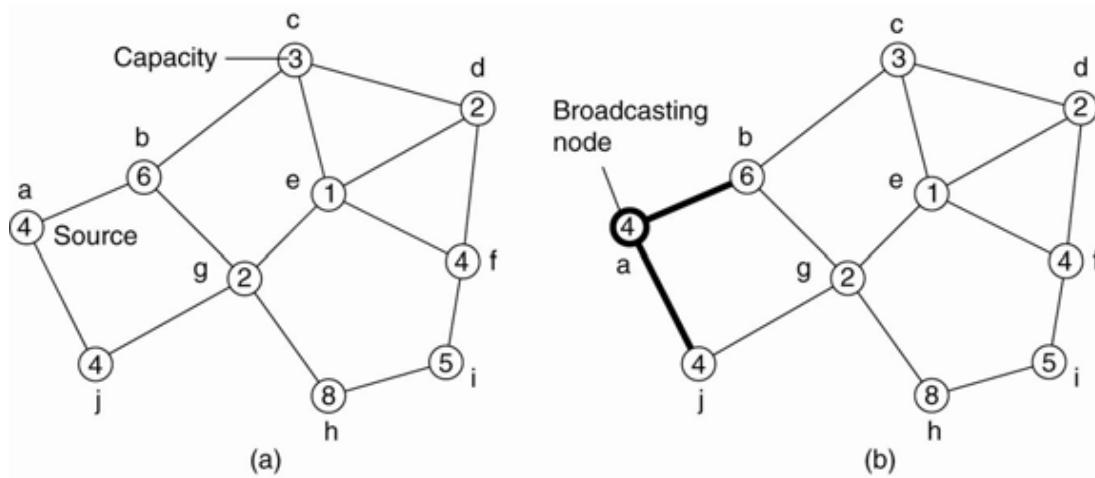
- Uložený soused je poté využijí pro zpětnou komunikaci. To znamená, že další komunikace už probíhá přes strom, nikoliv přes broadcast. Tím je ušetřena některé komunikace.

### 10.6.2 Složitost

Složitost z hlediska počtu zpráv.

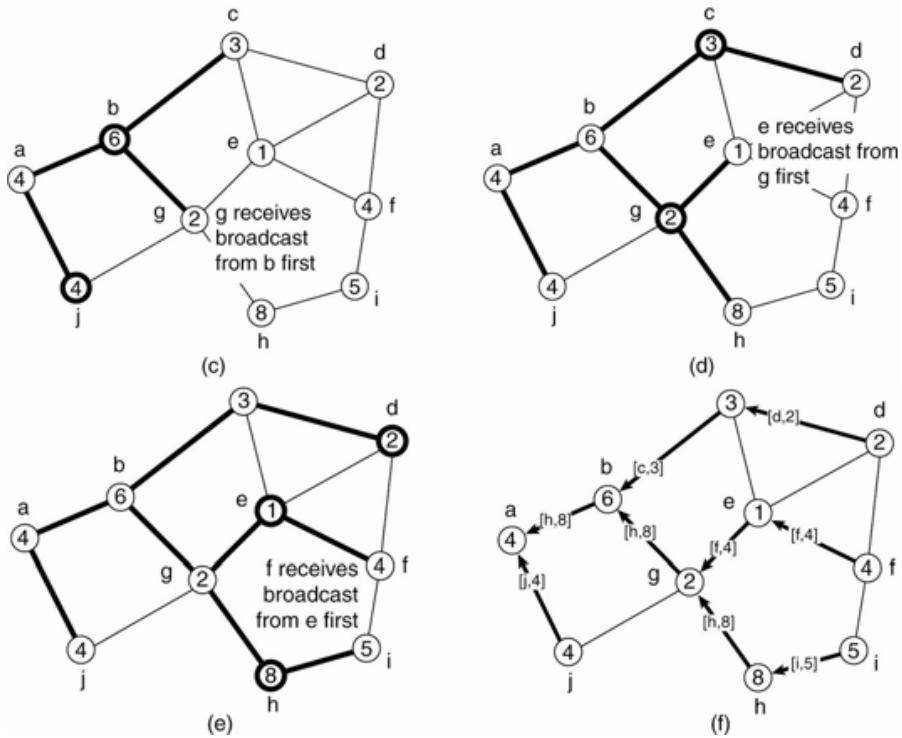
- Inicializační broadcast: počet hran grafu.
- Odpověď: počet hran kostry grafu.
- Result broadcast: počet hran kostry grafu.

### 10.6.3 Příklad



Node *a* initiates an election.

Obrázek 10.3: Příklad činnosti algoritmu pro obecnou topologii, část 1.



In the end, source *a* notes that *h* is the best leader and broadcasts this info to all nodes.

Obrázek 10.4: Příklad činnosti algoritmu pro obecnou topologii, část 2.

## Kapitola 11

# PDI – Podmínky konsistentního globálního stavu distribuovaného systému.

### 11.0.1 Zdroje

- 04\_Global\_State\_and\_Snapshot.pdf
- PDI\_2020-10-12.mp4

## 11.1 Úvod a kontext

**Distribuovaný systém** Distribuovaný systém je množina procesů  $p_1, p_2, \dots, p_n$ , které jsou propojeny komunikačními kanály. V systému neexistuje žádná globální paměť ani globální hodiny. Procesy spolu komunikují pouze zasláním zpráv skrze komunikačními kanály.

**Komunikační kanál** Komunikační kanál mezi procesy  $p_i$  a  $p_j$  značíme  $C_{ij}$ .

**Událost** Rozlišujeme tři typy událostí: interní událost procesu, zaslání zprávy a přijetí zprávy.

**Zpráva** Zpráva  $m_{ij}$  značí zprávu zaslanou procesem  $p_i$  procesu  $p_j$ .  $send(m_{ij})$  značí odeslání zprávy a  $recv(m_{ij})$  přijetí.

**Stav procesu** Lokální stav procesu  $p_i$  značíme  $LS_i$ . Lokální stav je definován jako sekvence všech událostí, o kterých proces  $p_i$  ví. Nechť  $e$  je libovolná událost,  $e \in LS_i$  značí, že událost  $e$  patří do lokálního stavu procesu  $p_i$ ,  $e \notin LS_i$  značí, že událost  $e$  nepatří do lokálního stavu procesu  $p_i$ .

**Stav komunikačního kanálu** Stav komunikačního kanálu  $C_{ij}$  značíme  $SC_{ij}$  a je definován množinou zpráv, které obsahuje. Pro kanál  $C_{ij}$  můžeme definovat jeho stav na základě lokálních stavů procesů  $LS_i$  a  $LS_j$ :

$$transit(LS_i, LS_j) = \{m_{ij} \mid send(m_{ij}) \in LS_i \wedge rec(m_{ij}) \notin LS_j\}$$

## 11.2 Model komunikace

- FIFO – Komukační kanál funguje jako fronta zpráv *first in, first out*. Kanál tedy zachovává pořadí zpráv sám o sobě.
- non-FIFO – Komunikační kanál se chová jako datová struktura množina, do které odesílatel vkládá zprávy a příjemce je odebírá v náhodném pořadí.
- Causal ordering (kauzální uspořádání) – Systém, který podporuje kauzální doručení zpráv splňuje následující vlastnost. Pro jakékoliv dvě zprávy  $m_{ij}$  a  $m_{kj}$  platí, pokud  $send(m_{ij}) \rightarrow send(m_{kj})$ , pak i  $recv(m_{ij}) \rightarrow recv(m_{kj})$ .

## 11.3 Konzistentní globální stav

**Globální stav** Globální stav distribuovaného systému je kolekce lokálních stavů procesů a komunikačních kanálů.

$$GS = \left\{ \bigcup_i LS_i, \bigcup_{i,j} SC_{ij} \right\}$$

**Časoprostorový diagram** Diagram pro vizualizaci komunikace procesů v distribuovaném systému. Viz obrázek 11.1 a 11.2.

**Konzistentní globální stav** Konzistentní globální stav (*snapshot*) je stav systému v určitém časovém okamžiku. Lze si jej představit jako řez v časoprostorovém diagramu, který rozděluje diagram na dvě části: minulost a budoucnost. Aby byl řez (globální stav) konzistentní, tak pokud je doručení nějaké zprávy v minulosti, musí být v minulosti i její odeslání. Formálně jde o globální stav, který splňuje následující podmínky:

$$send(m_{ij}) \in LS_i \Rightarrow m_{ij} \in SC_{ij} \oplus recv(m_{ij}) \in LS_j$$

,

$$send(m_{ij}) \notin LS_i \Rightarrow m_{ij} \notin SC_{ij} \wedge recv(m_{ij}) \notin LS_j$$

**K čemu je *snapshot*** *Snapshot* lze využít např. pro tvorbu záloh systému nebo při zotavování systému po chybách.

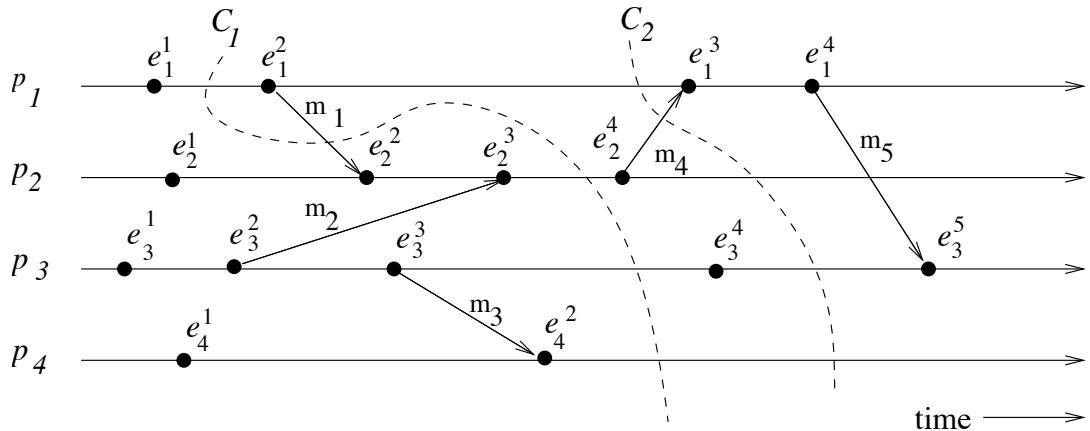
**Jak lze *snapshot* vytvořit** Absence globální sdílené paměti, globálních hodin a nepředvídatelná délka zpoždění v odesílání zpráv v distribuovaném systému činí problém vytváření snapshotů netriviálním. Způsob vytváření lze rozdělit do dvou kategorií: na základě algoritmů a na základě checkpointů.

**Problémy při zaznamenávání *snapshotu*** Jak rozlišit mezi zprávami, které mají být součástí snapshotu a které nikoliv?

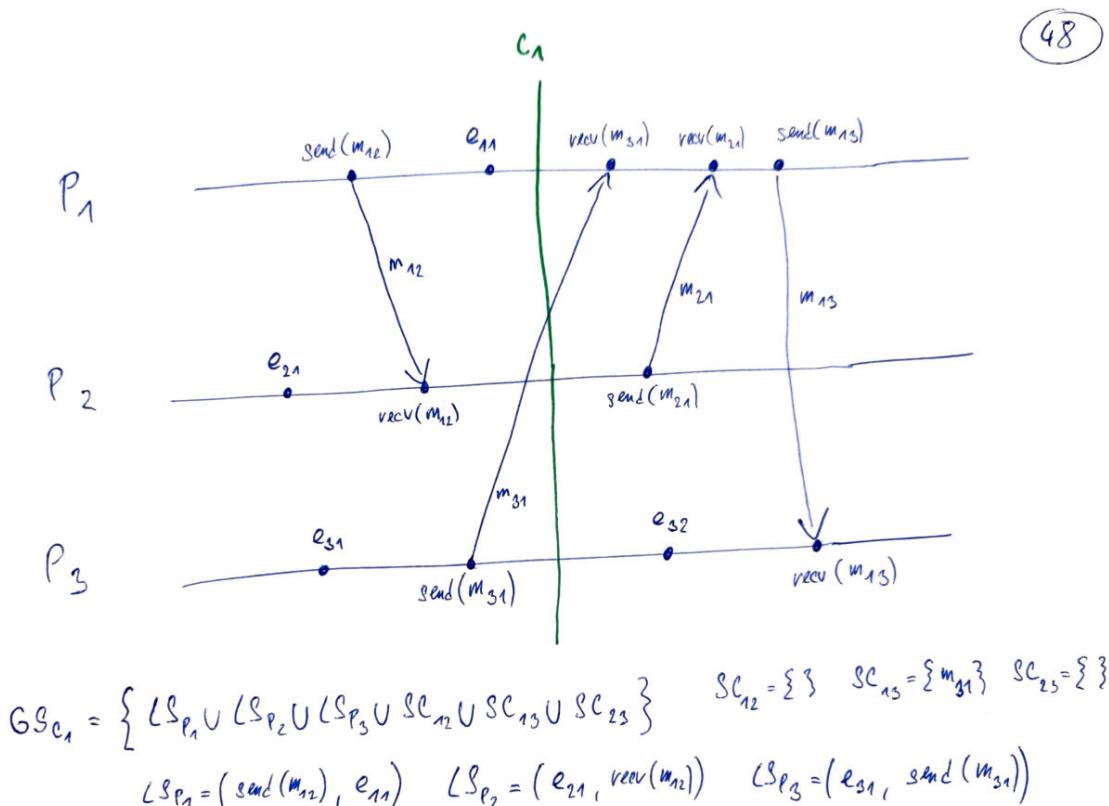
- Zprávy, které jsou odeslány procesem před zaznamenáním svého snapshotu, jsou zaznamenány do stavu.
- Zprávy, které jsou odeslány procesem po zaznamenání svého snapshotu, nejsou zaznamenány do stavu.

Jak rozpozнат okamžik, ve kterém má proces zaznamenat snapshot?

- Proces  $p_j$  musí zaznamenat svůj snapshot před zpracováním zprávy  $m_{ij}$ , která byla poslána procesem  $p_i$  po zaznamenání jeho snapshotu.



Obrázek 11.1: Příklad řezu v časoprostorovém diagramu. Řez  $C_1$  je nekonzistentní, kvůli zprávě  $m_1$ . Řez  $C_2$  je konzistentní a zpráva  $m_4$  je zachycena ve stavu kanálu  $Ch_{21}$ .



Obrázek 11.2: Příklad konzistentního globální stavu formálně.

## Kapitola 12

# PDI – Principy distribuovaného zpracování MapReduce, průběh a jednotlivé operace distribuovaného výpočtu pomocí MapReduce, jeho implementace v Apache Hadoop a Apache Spark.

### 12.1 Zdroje

- 09\_Hadoop.pdf
- 10\_Spark.pdf
- PDI\_2020-11-16.mp4
- PDI\_2020-11-23.mp4

### 12.2 Úvod a kontext

**OLTP** OLTP (*Online Transactional Processing*, provozní databáze, systémy pro online zpracování transakcí) jsou standardní databázové systémy s pevnou strukturou dat definovou pomocí databázového schématu. Jsou navrženy a optimalizovány pro chod provozních aplikací s primáním cílem zajistit rychlý a souběžný přístup k datům. To vyžaduje transakční zpracování, řízení souběžnosti a techniky obnovy (rollback), které zaručují konzistence dat. Díky této vlastnosti mají OLTP databáze špatný výkon při provádění složitých dotazů, které potřebují spojit mnoho relačních tabulek dohromady nebo agregovat velké objemy dat. Kromě toho obsahují typicky podrobná data a neobsahují historická data, která jsou při datové analýze potřeba.

**OLAP** OLAP (*Online Analytical Processing*, online analytické zpracování) je databázové paradigma specificky zaměřené na dotazy, zejména na analytické dotazy. Používají se zde jiné techniky indexování a optimalizace dotazů. Normalizace není pro toto paradigm žádoucí, protože rozděluje databázi na mnoho tabulek. Složité dotazy v takovém případě vyžadují rekonstrukci dat a s tím spojený vysoký počet spojování tabulek. Pracuje se s tzv.

multidimensionálními kostkami, avšak v pozadí jsou stále relační databáze.

**NoSQL** Potřeba ukládat proudy dat (zpracovávané v reálném čase bez možnosti poza-stavení), obrázky, multimédia, velké JSON soubory, ..., vedla ke vzniku NoSQL databází. NoSQL databáze používají jiné prostředky než tabulková schémata tradiční relační data-báze. Často jde o „hloupé“, nestrukturované uložiště klíč-hodnota.

**BigData** Velká, nestrukturovaná (různorodá), rychle rostoucí data, která není možné uložit ani zpracovávat běžnými přístupy (na jednom uzlu, jedním uzlem). Produkují je např.: IoT senzory, sociální sítě, chatovací aplikace, webové vyhledávače, ... Pro jejich zpracování je nutné využít distribuované systémy (pro uložení i zpracování).

**Distribuované zpracování dat** Distribuované zpracování dat je zpracování velkých dat (*big data*) pomocí distribuovaných systémů. To s sebou přináší problémy. Jak zaručit vhodnou distribuci dat a výpočtu mezi uzly? Jak řešit nespolehlivost a výpadky uzlů? Jak a kam zajistit doručení výsledků výpočtu? ...

## 12.3 MapReduce

Algoritmy pro indexování webových stránek (Page Rank) přestávaly být udržitelné, bylo potřeba zvýšit jejich škálovatelnost. Google vydal příspěvek „MapReduce: Simplified Data Processing on Large Clusters“, kde bylo představeno paradigma MapReduce. Jde o paradigma distribuovaného výpočtu založené na funkích *map* a *reduce* z funkcionálního programování.

**Map** Funkce *map* má ve funkcionálním programování 2 vstupní parametry a vrací seznam hodnot. První parametr je unární operátor (nebo funkce fungující jako unární operátor) a druhý je seznam hodnot. Výstupní seznam je spočítán jako aplikace unárního operátoru na vstupní seznam. Příklad:

$$\text{map}(\text{square}, [1, 2, 3, 4]) = [1, 4, 9, 16]$$

- . V paradigmata MapReduce *map* vrací data jako seznam dvojic klíč-hodnota, přesněji:

$map((key, value)) \rightarrow [(key, value)]$

**Reduce** Funkce *reduce* má ve funkcionálním programování 2 vstupní parametry a vrací jednu hodnotu. První parametr je binární operátor (nebo funkce fungující jako binární operátor) a druhý je seznam hodnot. Výstupní hodnota je spočítána jako postupná aplikace binárního operátoru na všechny hodnoty ve vstupním seznamu. Příklad:

*reduce*(+, [1, 4, 9, 16]) = 30

- . V paradigmata MapReduce *reduce* bere na vstupu klíč a seznam hodnot a vrací opět seznam dvojic klíč-hodnota, přesněji:

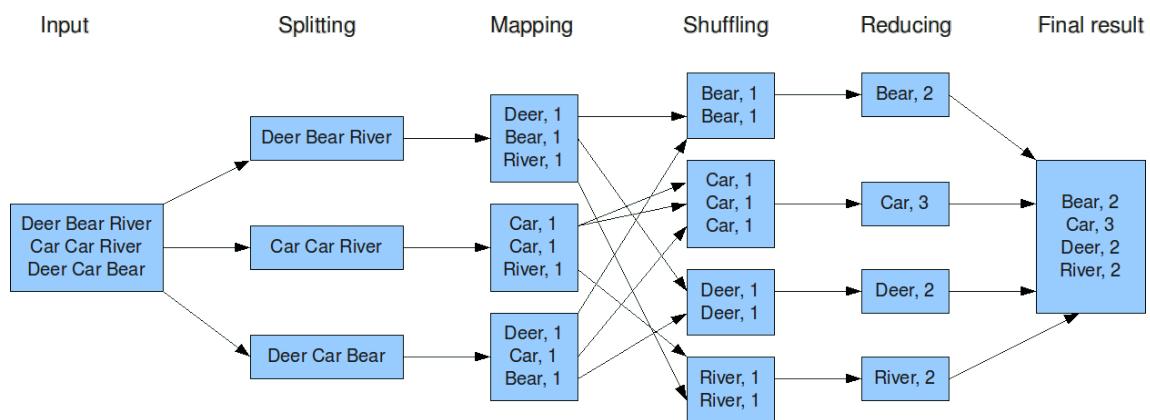
$reduce(key, [value]) \rightarrow [(key, value)]$

```

1 def map(input_key: str, input_value: str) -> list[tuple[str, int]]:
2     # input_key - document name
3     # input_value - document content (etc. line)
4     result = []
5     for word in input_value.split(' '):
6         result.append((word, 1))
7     return result
8
9 def reduce(input_key: str, input_value: list[int]) -> tuple[str, int]:
10    result = 0
11    for val in input_value:
12        result += value
13    return (input_key, result)

```

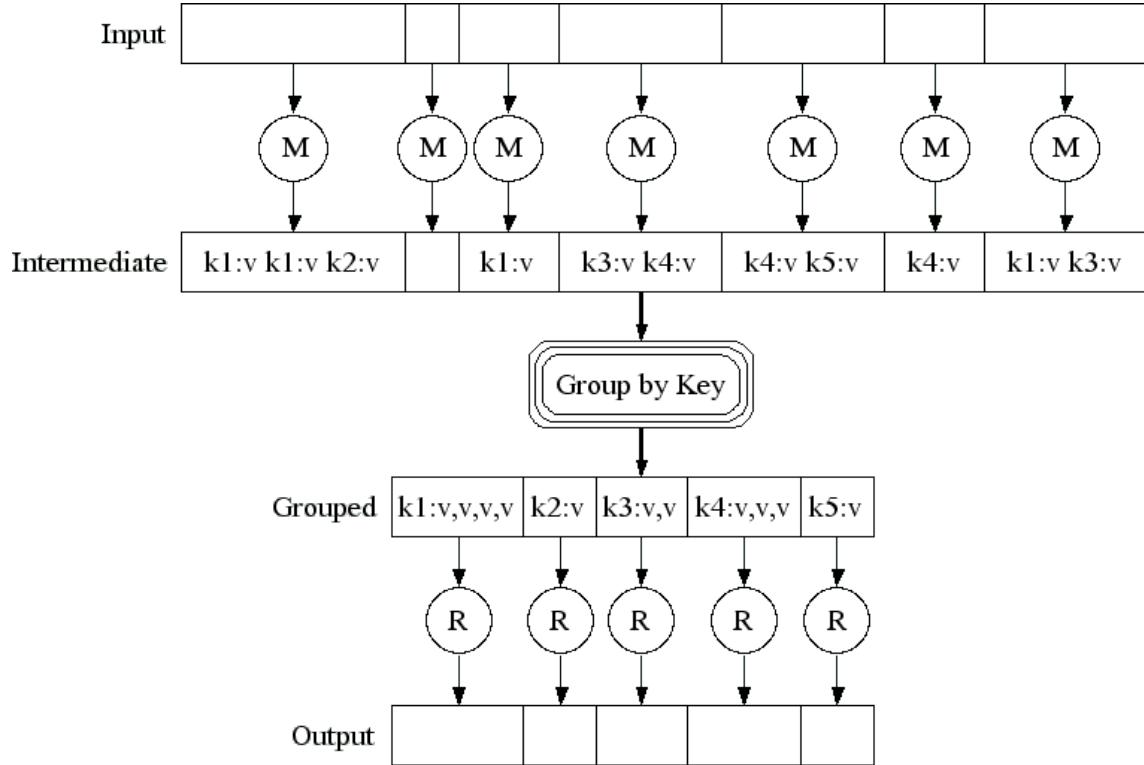
Výpis 12.1: Příklad implementace funkcí *map* a *reduce* v paradigmatu MapReduce pro počítání četnosti slov ve vstupu v Pythonu.



Obrázek 12.1: Úloha počítání četnosti slov v paradigmatu MapReduce v diagramu.

**Průběh MapReduce** Celý MapReduce probíhá v několika krocích, viz obrázek 12.1.

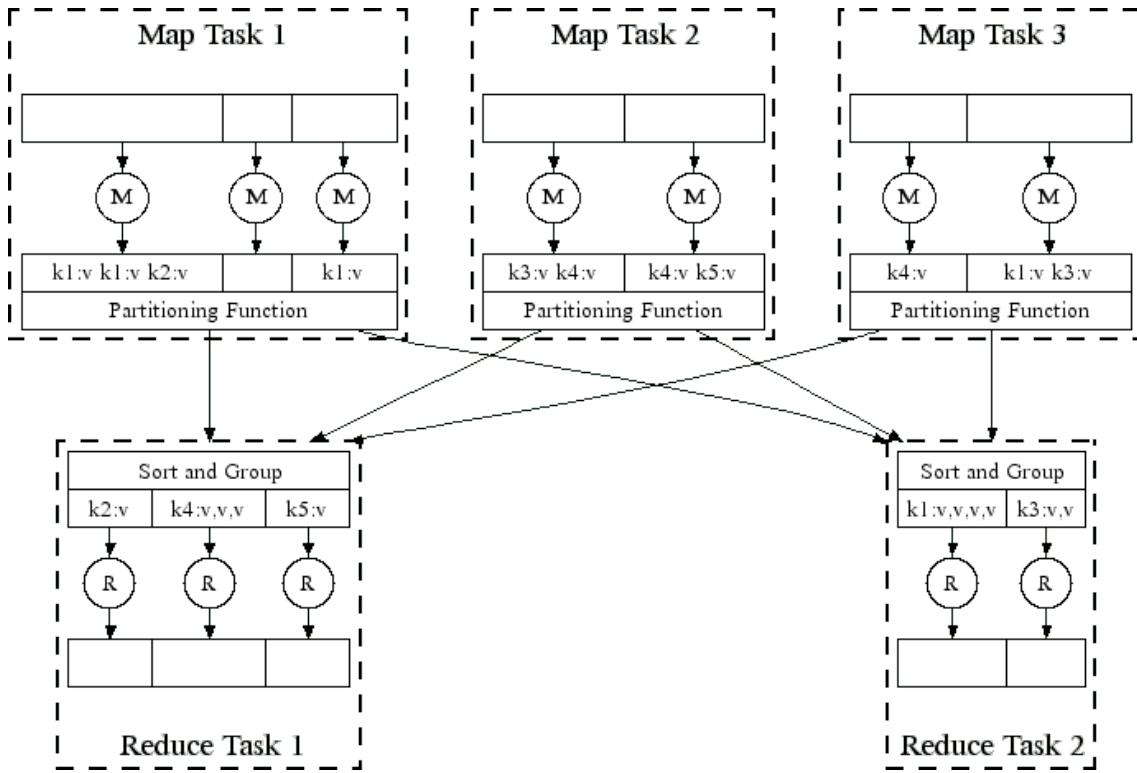
1. Input – Přípravený vstup pro distribuovaný výpočet (např. soubory ve virtuálním distribuovaném souborovém systému, viz dále HDFS).
2. Splitting – Rozdělení vstupu na části, které budou přiděleny jednotlivým uzlům. Může být výchozí (např. rozdělení textového souboru po řádcích) nebo definováno uživatelem.
3. Mapping – Každý uzel aplikuje funkci *map* na svoji přidělenou část. Uživatel definuje jak má funkce *map* vypadat.
4. Shuffling (také Grouping, Partitioning, Comparing) – Výpočetní uzly si mezi sebou vyměňí hodnoty, které spočítaly, na základě klíče. Tento krok zajišťuje platforma pro distribuovaný výpočet sama o sobě, typicky na základě hashů klíčů. Tento krok je většinou *bottleneck*.
5. Reducing – Každý uzel zapojený do tohoto kroku (často je v tomto kroku potřeba méně uzlů, než v kroku mapping) aplikuje funkci *reduce* na svoji přidělenou část. Uživatel definuje jak má funkce *reduce* vypadat.
6. Final Result – Finální výsledek (např. zapsán do do virtuálního distribuovaného souborového systému, viz dále HDFS).



Obrázek 12.2: Výpočet MapReduce v obecném schématu.

**Combiner** Optimalizační krok, jde o „jakési“ provedení operace *reduce* už ve fázi *map* (každým uzlem). Tím je snížen počet mezivýsledků ve fázi Shuffling. Typicky funkce *combine* je stejná jako *reduce*.

**Virtuální distribuovaný souborový systém** Pro realizaci distribuovaného výpočtu je rovněž potřeba distribuovaný souborový systém (DFS). Ten je typicky realizován jako virtuální souborový systém nad jednotlivými souborovými systémy uzlů. Např.: GFS – Google File System, HDFS – Hadoop File System (viz dále). DFS obsahuje data samotná (*data nodes*) a metadata o tom, která data jsou na jakých uzlech (*name nodes*).



Obrázek 12.3: Výpočet MapReduce v obecném schématu a rozdělením práce na jednotlivé uzly (uzel je typicky víceprocesorový).

## 12.4 Apache Hadoop

Apache Hadoop je *open-source* implementace MapReduce paradigmatu vyvíjená Apache Software Foundation. Jde o implementaci v Java, ta je vhodná, jelikož díky JVM (Java Virtual Machine) je spouštění uživateli definovaných funkcí *map* a *reduce* snadné.

**Hadoop MapReduce** – Implementace MapReduce paradigmata. Data jsou čtena a ukládána na HDFS (včetně mezivýsledků). To znamená, můžeme pracovat v podstatě neomezenými daty, ale ukládání a načítání výpočet zpomalují.<sup>1</sup>

**HDFS** HDFS (*Hadoop Distribute File System*) je virtuální distribuovaný souborový systém. Standardní soubor je rozdělen na datové bloky které jsou distribuovány na různé datové uzly. Architektura HDFS se skládá ze dvou typů uzlů – Name Node a Data Node. Name Node obsahuje alokační tabulkou pro souborový systém. Ví které datové bloky patří kterému souboru a kde jsou uloženy. Obsahuje další metadata jako názvy souborů, cesty, ... Data Node obsahuje datové bloky. Typicky redundancy a replikace, počítá se s možným selháním uzlů. Pro **čtení dat** se klient zeptá Name Nodu na konkrétní soubor v HDFS. Name Node vrátí metadata o souboru, na jakých Data Nodech se vyskytuje. Klient požádá příslušné Data Nody, ty mu pošlou data, která se na klientovi „poskládají“ do výsledného souboru. Pro **zápis dat** se klient zeptá Name Nodu, kam by měl zapisovat. Klient zapíše na příslušný Data Node. Data Node poté vyřeší replikace s dalšími uzly.

<sup>1</sup> Nebylo přednášeno podrobněji, pravděpodobně stačí princip obecného MapReduce, který byl vysvětlen v předchozí sekci.

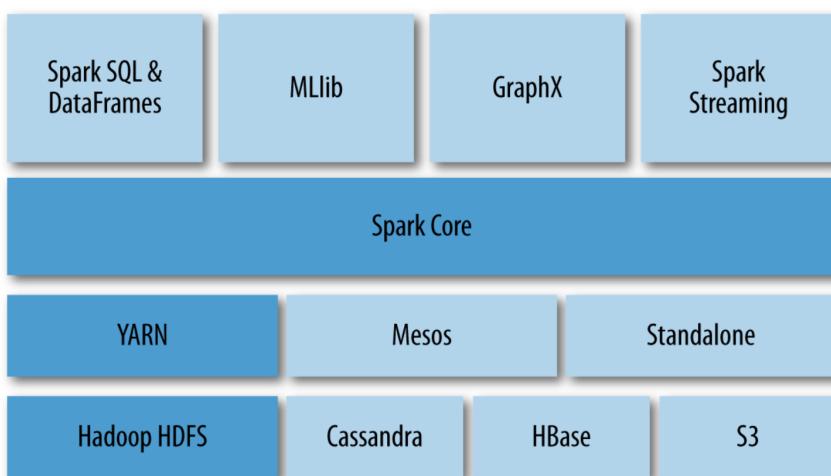
**Hadoop YARN** Hadoop YARN je plánovač (*scheduler*). Plánuje výpočet tak, aby proběhl co nejlepším způsobem na konkrétní distribuované architektuře. Plánovač má obecné obecné rozhraní a Hadoop YARN lze nahradit za jiný.

**Hadoop Common** Hadoop Common jsou další knihovny a ovladače pro klienty.

**Další nástroje** Nad Apache Hadoop existuje mnoho dalších nástrojů. Apache Pig pro *high level* programování map-reduce úloh. Apache Hive pro dolování dat nad Apache Hadoop. Apache HBase jako distribuovaná databáze nad Apache Hadoop, ...

## 12.5 Apache Spark

Apache Spark je *open-source* nástroj pro distribuované zpracování rozsáhlých dat vyvíjený Apache Software Foundation. Hlavní cíl je zvýšení rychlosti. Spark na to jde přesunutím co nejvíce výpočtů do operační paměti jednotlivých uzlů a tím pádem zminimalizovat počet zápisů a čtení z DFS (snaha odstranit *bottleneck* v kroku shuffling u Hadoopu). Tím ale vzniká jiný problém, a sice výpadek uzlu znamená, že data jsou ztraceny.



Obrázek 12.4: Architektura Apache Spark. Hlavní je Spark Core, zbytek funguje na systému pluginů a může používat HDFS, Hadoop YARN a Hadoop Common.

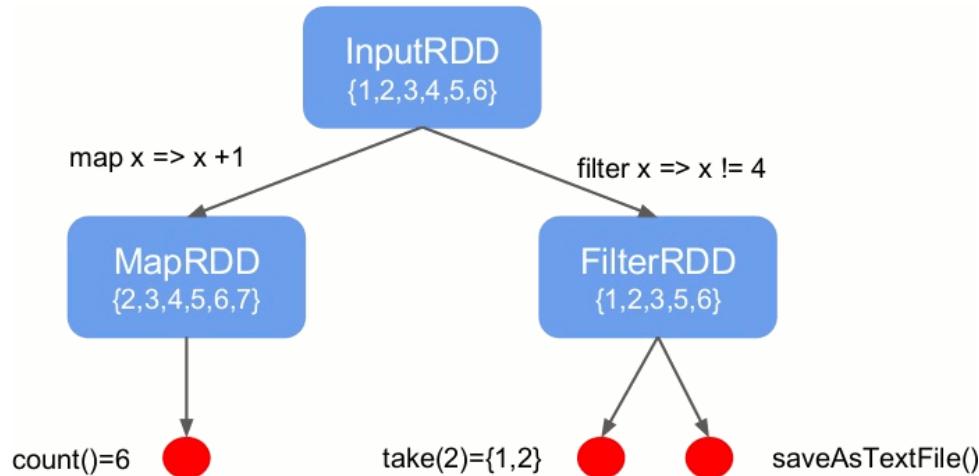
**Resilient Distributed Dataset** Resilient Distributed Dataset (RDD) je základní datová struktura Sparku. Jedná se o typované kolekce n-tic, které jsou neměnné (*read only*). Vstup je transformován na RDD a každá operace je pak transformace jednoho RDD na jiné.

$$RDD_1 \rightarrow map() \rightarrow RDD_2 \rightarrow reduce() \rightarrow RDD_3$$

**Strategie vyhodnocování** Spark uplatňuje strategii vyhodnocování *lazy evaluation*. Vyhodnocování výrazu je odkládáno až do doby, dokud není potřeba jeho hodnota. Zabraňuje opakovanému vyhodnocování. Je vyhodnocována pouze ta část, která je potřeba. RDD funguje jako abstraktní datová struktura, nemusí obsahovat data uvnitř, ale pouze předpis jak data získat a získá je, až když jsou potřeba.

**Struktura výpočtu** Struktura výpočtu odpovídá orientovanému acyklickému grafu (DAG, *Directed Acyclic Graph*). Uzly jsou RDD a hrany jsou transformace. DAG je znám i dalším uzlům, takže pokud nastane výpadek uzlu a výpočet je ztracen, může být uzel snadno zastoupen.

**Klíčové vlastnosti** Klíčové vlastnosti Sparku jsou *lazy evaluation*, *in-memory* a *parallel computing*.



Obrázek 12.5: Příklad výpočtu v Apache Spark.

# Kapitola 13

**KRY – Symetrická kryptografie.  
Vlastnosti, vlastnosti bezpečného  
algoritmu, délka klíče, útok silou,  
příklady symetrických algoritmů,  
Feistelovy šifry, DES, režimy činnosti,  
proudové šifry.**

## 13.1 Zdroje

- KRY03\_Sym\_MNG.pdf
- KRY\_2021-02-22.mp4
- KRY\_2021-03-01.mp4
- KRY\_2021-03-08.mp4

## 13.2 Úvod a kontext

**Kryptografie** Kryptografie (šifrování) je věda o metodách utajování smyslu zpráv převodem do podoby, která je čitelná jen se speciální znalostí.

**Kryptoanalýza** Kryptoanalýza je věda zabývající se metodami získávání obsahu šifrovaných informací bez přístupu k tajným informacím, které jsou za normálních okolností potřeba, tzn. především k tajnému klíči.

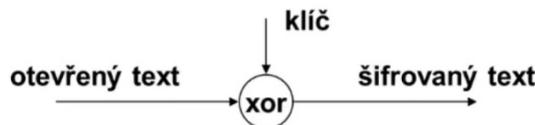
**Kryptologie** Jeden výraz pro kryptografii a kryptoanalýzu.

**Caesarova šifra** Princip Caesarovy šifry je založen na tom, že všechna písmena zprávy jsou během šifrování zaměněna za písmeno, které se abecedně nachází o pevně určený počet míst dále (tj. posun je pevně zvolen). Caesarova šifra spadá do kategorie substitučních šifer (stejný znak je při více vyskytech vždy zašifrován na stejný znak).

**Vigenerova šifra** Rozšíření Caesarovy šifry, klíč je delší než 1 znak. Klíč je řetězec, který reprezentuje posuny. V případě že vstup je delší než klíč, je klíč perioricky opakován. Vigenerova šifra spadá do kategorie polyalfabetických substitučních šifer (stejný znak může být při více výskyttech zašifrován na jiný znak).

**Vernamova šifra (*One Time Pad*)** Vernamova šifra spadá do kategorie polyalfabetických substitučních šifer a je i dnes nerozluštitevná pokud:

- klíč je delší než vstupní text,
- klíč se nepoužije opakováně,
- klíč je náhodný.



Obrázek 13.1: Vernamova šifra.

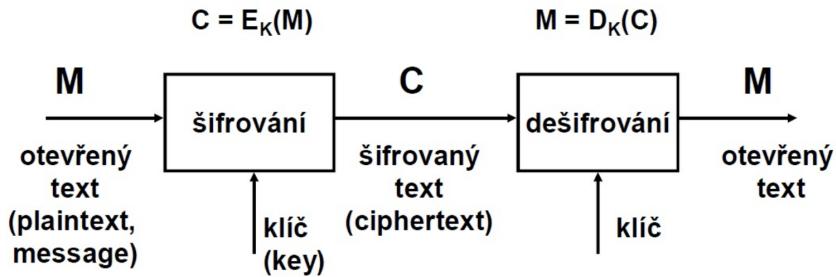
**Autoklíč (autokey)** Šifrování klíčem a když vstupní text je delší než klíč, tak se pokračuje šifrováním otevřeným nebo šifrovaným textem. Lze použít u Vigenerovy nebo Vernamovy šifry.

**Symetrická kryptografie** Algoritmy používají k šifrování i dešifrování stejný klíč. Výhodou symetrických šifer je jejich nízká výpočetní náročnost. Asymetrické šifry mohou být i stotisíckrát pomalejší. Nevýhodou je nutnost sdílení tajného klíče, takže jedna strana musí klíč vygenerovat a potom ho bezpečným způsobem předat druhé straně.

### Typy útoků

- Ciphertext Only Attack (COA) – Útočník zná pouze zašifrovaný text a snaží se zjistit klíč nebo otevřený text. Nejčastější případ.
- Known Plaintext Attack (KPA) – Útočník zná zašifrovaný text a otevřený text a snaží se zjistit klíč.
- Chosen Plaintext Attack (CPA) – Útočník zná to co v KPA a navíc si text může zvolit.

**Útok silou** Při útoku silou (*brute force*) zkouší útočník všechny teoreticky možné klíče, dokud nenajde ten správný.



Obrázek 13.2: Princip kryptografie, podle typu klíčů dělíme na symetrickou (tajný klíč) a asymetrickou (veřejný klíč, soukromý klíč).

**Bezpečný algoritmus** V moderní kryptografii je nepřijatelné utajování algoritmů (*security by obscurity*) – předpokládáme, že útočník zná šifrovací algoritmus. Bezpečnost musí záviset pouze na utajení klíče (Kerckhoffuv princip, *security by design*). Symetrický algoritmus je považován za bezpečný, pokud neexistuje rychlejší útok než útok silou.

**Délka klíče** Dnes je považováno 80 bitů a více za dostatečné. Typicky se délka zaokrouhuje na mocninu 2 (typicky 128b). Klíče symetrických algoritmů jsou kratší než asymetrických. Konkrétně: DES – 56b, 3DES – 112, AES – variabilní.

**Využití** Symetrická kryptografie je vhodná pro šifrování většího objemu dat. Narozdíl od asymetrické, která je pro tento účel příliš pomalá. Proto např. HTTPS využívá asymetrickou kryptografií pro výměnu symetrických klíčů a poté symetrickou kryptografií pro šifrování provozu.

**Vlastnosti moderní kryptografie** Symetrická kryptografie zaručuje všechny následující, kromě nepopiratelnosti – více entit má k dispozici klíč.

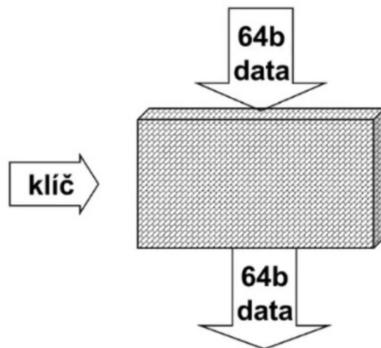
- Důvernost – Utajení informace. Bez znalosti klíče, není možné data číst.
- Autentizace – Prokázání, že zprávu skutečně poslal odesíatel a nikoliv útočník, který se za odesílatele vydává.
- Integrita – Prokázání, že nikdo nemohl data po cestě od odesílatele k příjemci změnit. Ochrana proti neoprávněné, neodhalené modifikaci zprávy.
- Nepopiratelnost – Pokud odesíatel data poslal, nemůže tuto skutečnost popřít.

### 13.3 Blokové šifry

Blokové šifry šifrují data po blocích pevně stanovené délky (64b, 128b, 256b, ...). Pokud je dat více, rozdělí se na více bloků, přičemž do zbylého místa v posledním je umístěno zarovnání *padding* (informace o délce zarovnání může být obsažena v posledním bytu). Příklady blokových šifer:

- Feistelova šifra (spíše princip)
- Data Encryption Standard (DES)
- Triple Data Encryption Algorithm (3DES)
- International Data Encryption Algorithm (IDEA)

- Blowfish
- Tiny Encryption Algorithm (TEA)
- Advanced Encryption Standard (AES)



Obrázek 13.3: Princip blokových šifer.

### 13.3.1 Feistelova šifra

Feistelova šifra (Feistelův princip) je koncept šifrování, který konkrétní algoritmy využívají. Jedná se o substituční-permutační síť. Vstupní blok je rozdělen na dvě poloviny  $L$  a  $R$ , výpočet výstupu pak vypadá následovně.

$$L_i = R_{i-1} \quad (13.1)$$

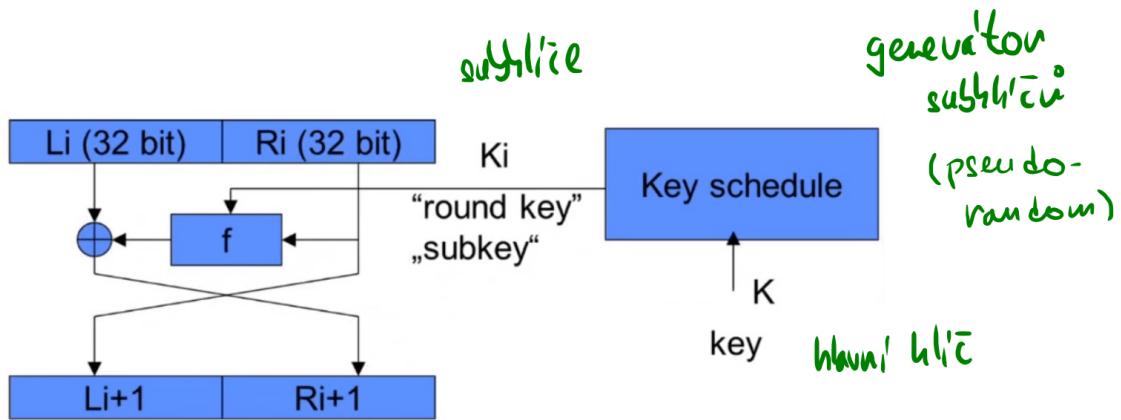
$$R_i = L_{i-1} \oplus F(R_{i-1}, K_i) \quad (13.2)$$

**Funkce  $F$**   $F$  je funkce, na kterou Feistelova šifra neklade žádné požadavky. Jednotlivé algoritmy, využívající Festelovu šifru, funkci samy definují. Požadavky na funkci  $F$ , aby algoritmus byl bezpečný:

- skrytí vlastností zprávy;
- skrytí vlastností zprávy.

**Subklíč**  $K$  je tzv. subklíč, který je generován typicky nějakým pseudonáhodným generátorem na základě inicializačního klíče (hlavní).

**Dešifrování** Dešifrování se provádí stejným způsobem, pouze pořadí subklíčů je opačné.



Obrázek 13.4: Jeden krok opakování (Feistelův krok) vizuálně.

## Příklad

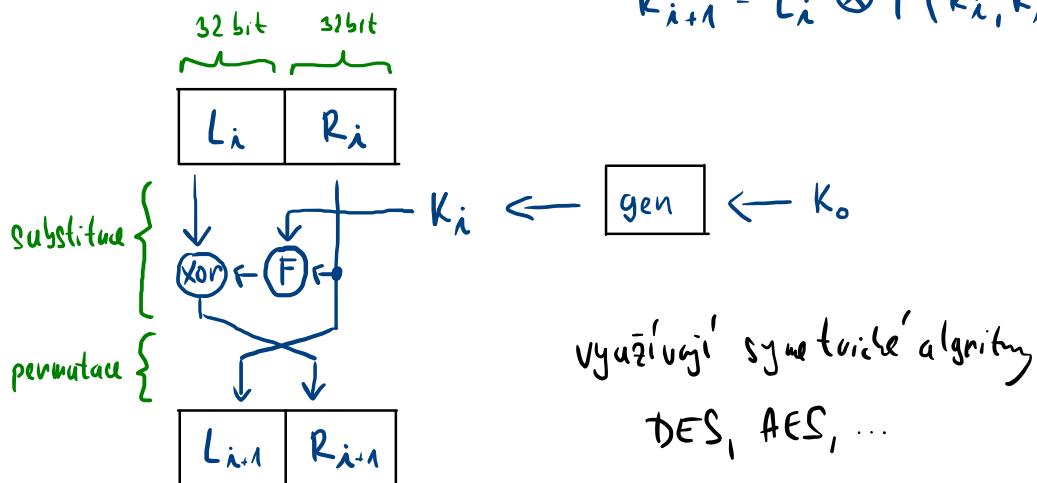
9. Nakreslit a popísať Feistelovu šifru, napísat algoritmus ktorý to používa.

- Princip klasických symetrických říšer

- Vstup rozdelen na bloky o 64 bitech  
↳ vstup rozšířen

$$L_{i+1} = R_i$$

$$R_{i+1} = L_i \otimes F(R_i, K_i)$$



Obrázek 13.5: Feistelova šifra – příklad a rekapitulace.

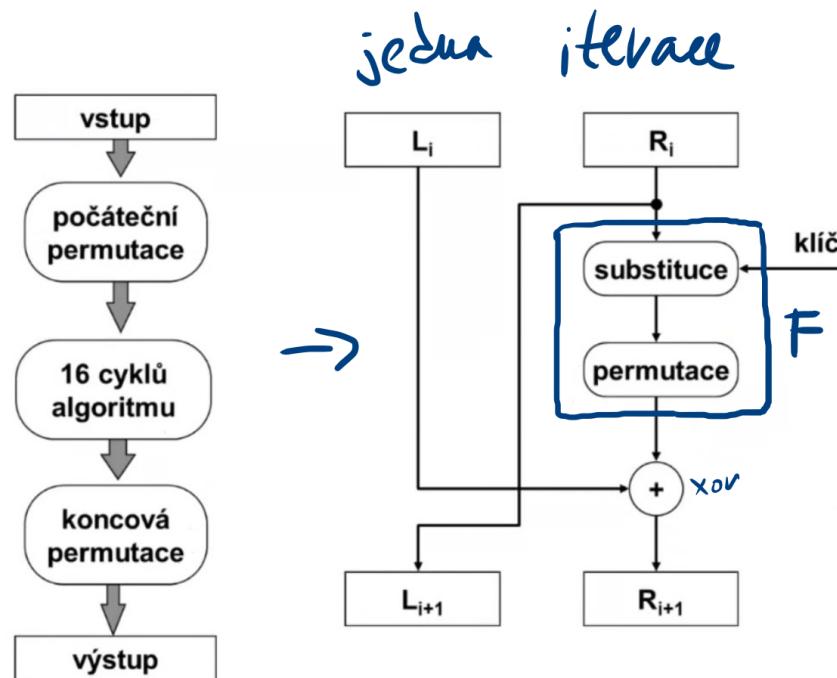
### 13.3.2 Data Encryption Standard (DES)

DES byl první algoritmus s veřejnou specifikací (*security by design*). Využívá princip Feistelovy šifry – 16 kol. Dodatečně přidává na začátek a konec permutaci navíc. Klíč je dlouhý 64b (resp. 56 významových bitů a 8 paritních).

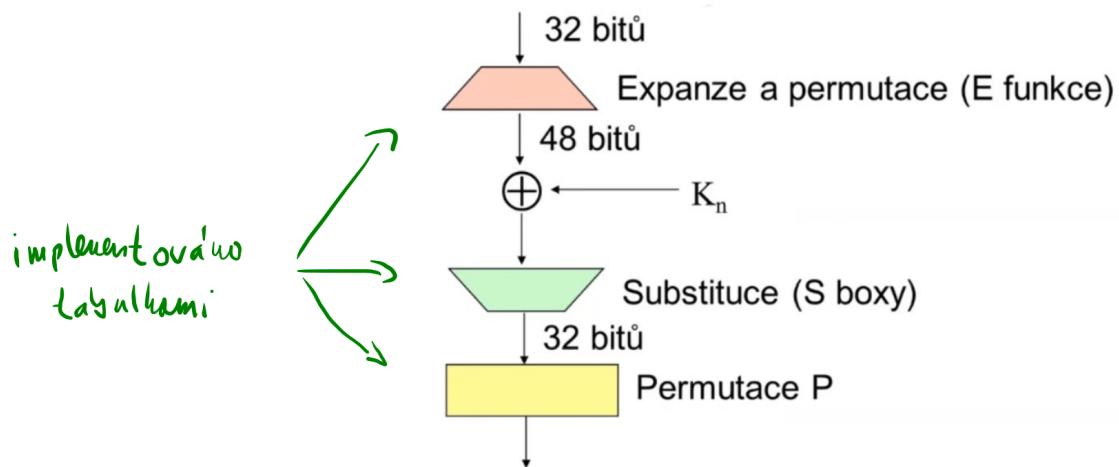
#### Slabiny

- 56 bitový klíč je příliš krátky a je možný útok silou.

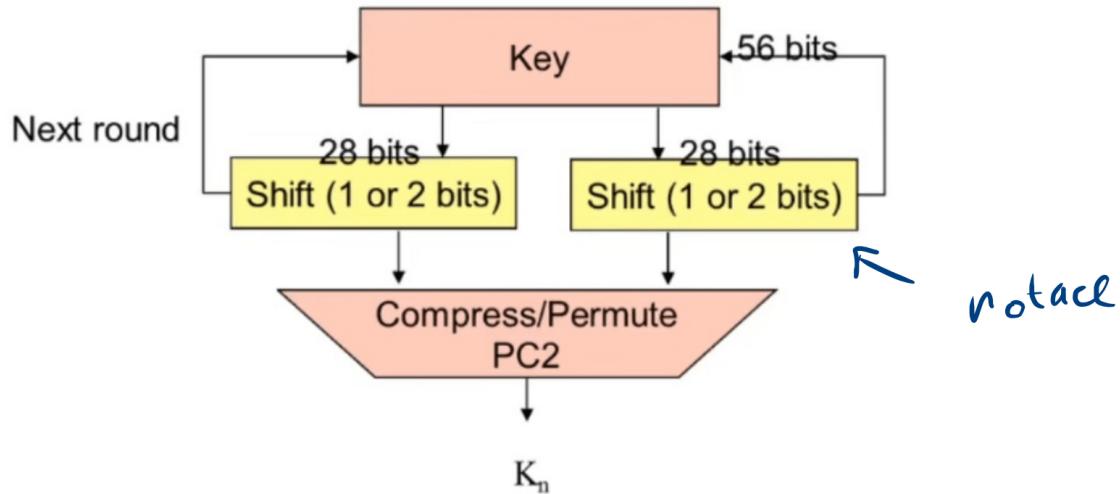
- Rozdílná velikost bloku a klíče (zvláštnost).
- Existence slabých a poloslabých klíčů.
- Není jasné proč zrovna 16 iterací a zda je to dostatečné.



Obrázek 13.6: DES – Schéma fungování algoritmu.



Obrázek 13.7: DES – funkce F.



Obrázek 13.8: DES – generování subklíčů.

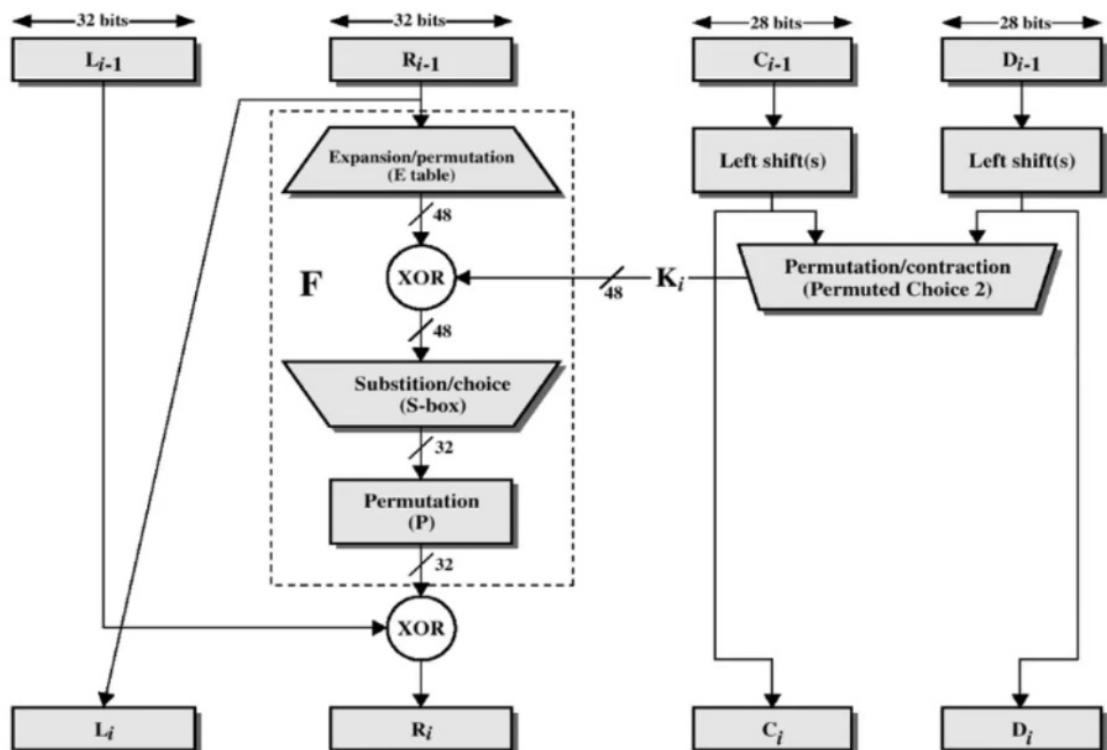
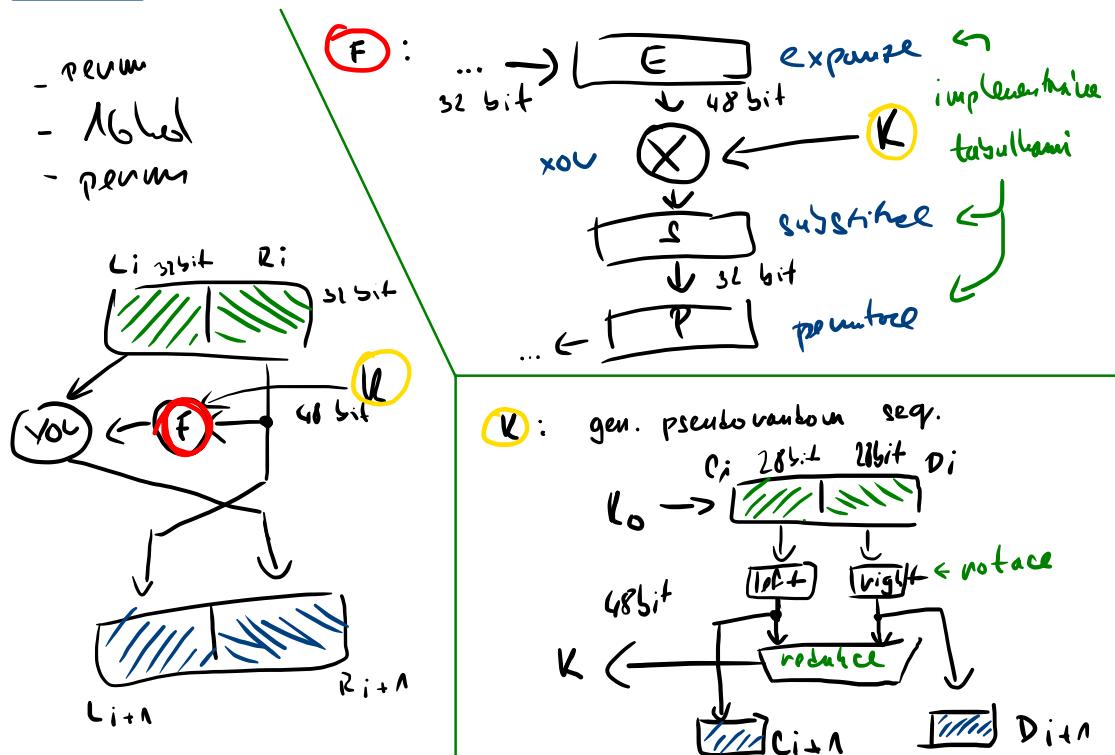


Figure 2.4 Single Round of DES Algorithm

Obrázek 13.9: DES – jedno kolo algoritmu. *Left shift* je ve skutečnosti bitová rotace.

## Příklad

3. Nakreslete schema DES, včetne key scheduler, popiste všechny 3 casti funkce F, jaký mají uzel a napiste proč je treba pochybovat o bezpečnosti DESu



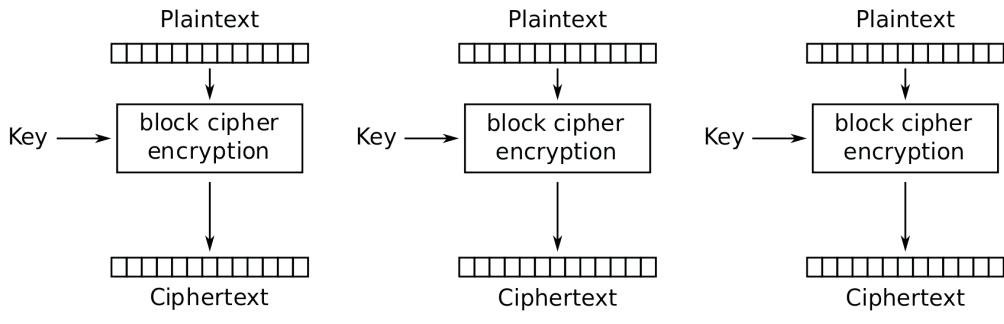
Obrázek 13.10: DES – příklad a rekapitulace.

### 13.4 Provozní režimy činnosti blokových šifer

Jak použít blokové šifry abychom byli schopni šifrovat data delší než jeden blok?

#### 13.4.1 Electronic Code Book (ECB)

ECB („kódová kniha“) je výchozí *naivní* režim. Bloková šifra se při něm přímo aplikuje nezávisle na jednotlivé bloky, tedy při daném klíči odpovídá stejnému bloku otevřeného textu stejný blok šifrového textu. To má nežádoucí důsledky z hlediska bezpečnosti, v datech zůstane původní struktura, např. šifrovaný obrázek je rozpoznatelný.



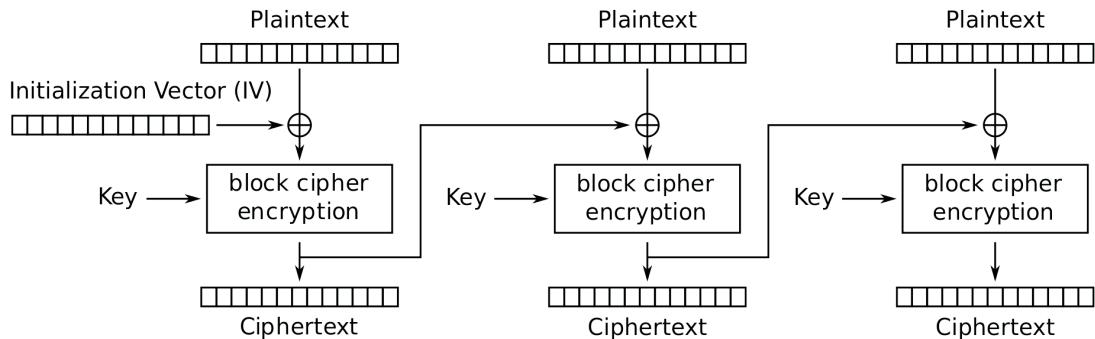
Electronic Codebook (ECB) mode encryption

Obrázek 13.11: Ukázka režimu ECB.

### 13.4.2 Cipher Block Chaining (CBC)

V režimu CBC („řetězení šifrových bloků“) je každý blok před šifrováním xorován zašifrovaným předchozím blokem a první blok je xorován inicializačním vektorem. Tento režim je široce používán. Nevýhody plynou ze zřetězené závislosti (šifrovaný blok závisí na všech předcházejících): Šifrování nelze paralelizovat a při poškození šifrového bloku nelze dešifrovat ani blok přímo následující. Dešifrování paralelizovat lze.

$$\begin{aligned} C_i &= E_K(P_i \oplus C_{i-1}) \\ P_i &= D_K(C_i) \oplus C_{i-1} \\ C_0 &= IV \end{aligned} \quad (13.3)$$



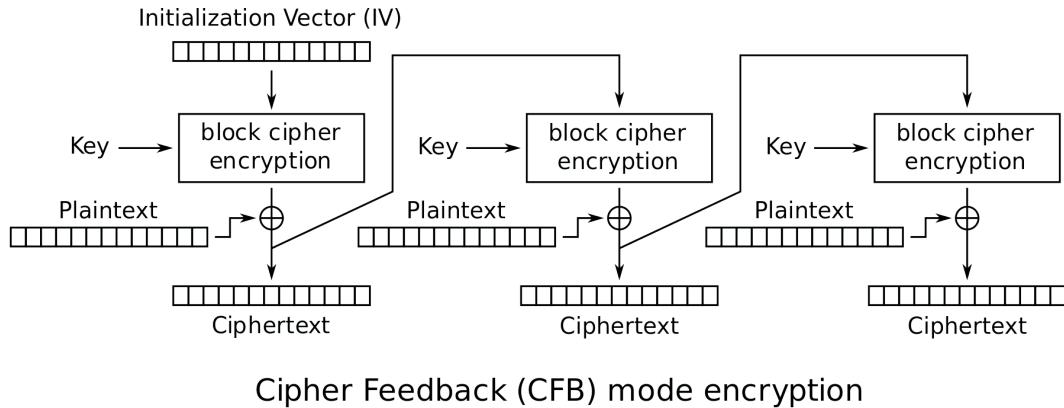
Cipher Block Chaining (CBC) mode encryption

Obrázek 13.12: Ukázka režimu CBC.

### 13.4.3 Cipher Feedback (CFB)

Režim CFB (šifrová zpětná vazba) se liší oproti CBC v prohození pořadí operací xor a šifrování – nejprve se zašifruje předchozí šifrovaný blok (resp. inicializační vektor) a výsledek se xoruje s otevřeným blokem. Toto prohození má významné implementační dopady: díky symetrii operace XOR vypadá dešifrovací funkce obdobně jako šifrovací. Šifruje pomaleji než CBC. Vstup není nutné zarovnávat. Plynou stejné nevýhody jako pro CBC.

$$\begin{aligned}
 C_i &= E_K(C_{i-1}) \oplus P_i \\
 P_i &= E_K(C_{i-1}) \oplus C_i \\
 C_0 &= IV
 \end{aligned} \tag{13.4}$$



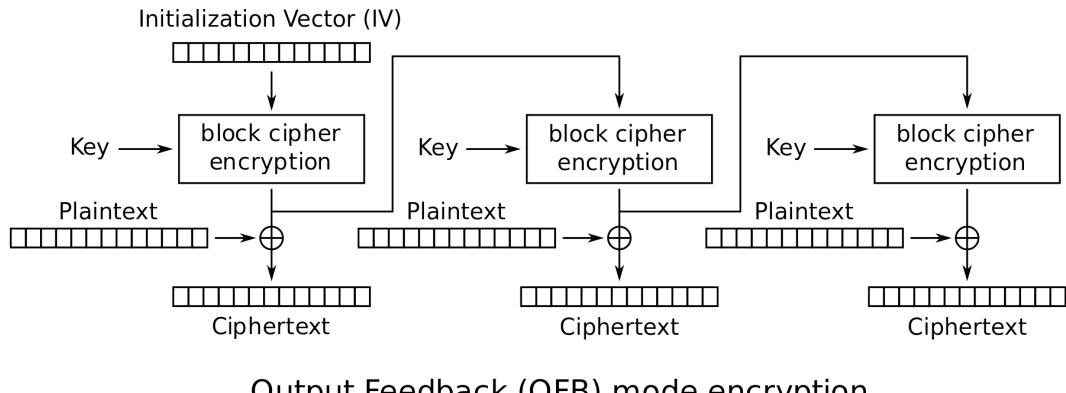
Obrázek 13.13: Ukázka režimu CFB.

#### 13.4.4 Output Feedback (OFB)

Režim OFB (*výstupní zpětná vazba*) se liší od CFB pouze v tom, kde bere zpětnou vazbu. Šifrování probíhá pouhým xorováním otevřeného bloku s heslem, které je v každém kroku zašifrováno použitou blokovou šifrou. První blok hesla je získán zašifrováním inicializačního vektoru. Režim převádí blokovou šifru na synchronní proudovou šifru.

**Slabina** Celý blok encryption je pouze generátor pseudonáhdon posloupnosti (je nezávislá na otevřeném nebo šifrovaném textu). To umožňuje Known Plaintext Attack. Z toho plyne, že jedním klíčem není bezpečné šifrovat více než jednu zprávu.

$$\begin{aligned}
 C_i &= E_K(C_{i-1}) \\
 P_i &= P_i \oplus C_i \\
 C_0 &= IV
 \end{aligned} \tag{13.5}$$



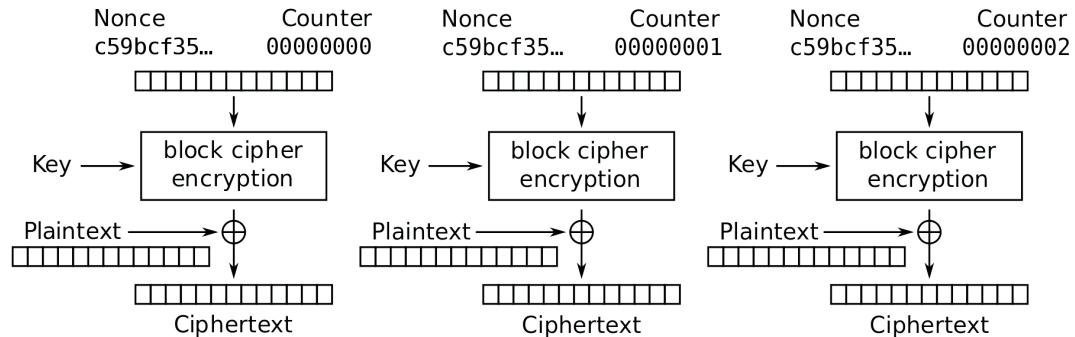
Output Feedback (OFB) mode encryption

Obrázek 13.14: Ukázka režimu OFB.

### 13.4.5 Counter (CTR)

Režim CTR („čítačový režim“) převádí stejně jako OFB blokovou šifru na synchronní proudovou. Heslo, se kterým se blok otevřeného textu xoruje, je však získáno zašifrováním čítače, který se každou iteraci zvětšuje o pevně danou hodnotu, zpravidla o 1. Obsah čítače je opět před šifrováním nastaven inicializačním vektorem. Každý blok je šifrován nezávisle na ostatních, díky tomu je možné paralelizovat.

$$\begin{aligned} CTR_i &= CTR_{i-1} + 1 \\ P_i &= P_i \oplus E_k(CTR_i) \\ CTR_0 &= IV \end{aligned} \quad (13.6)$$



Counter (CTR) mode encryption

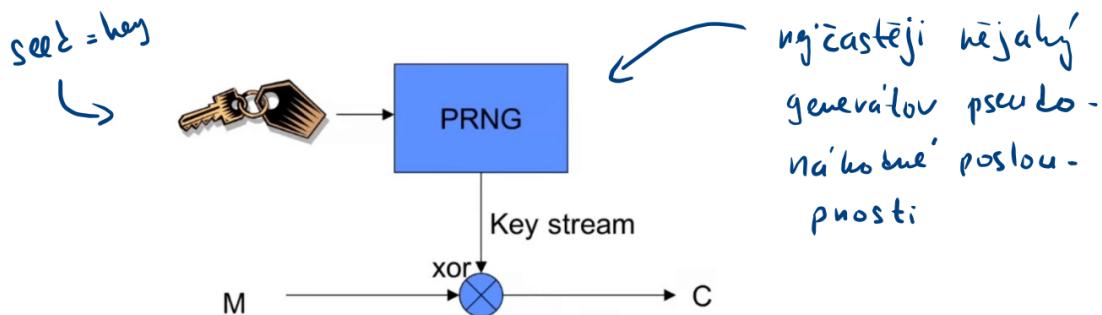
Obrázek 13.15: Ukázka režimu CTR.

## 13.5 Proudové šifry

Proudové šifry šifrují data jako *proud* (stream), nejčastěji po jednotlivých bytech. Dešifrování vždy probíhá stejným způsobem. Proudové šifry jsou rychlejší než blokové šifry a pro implementaci potřebují jednodušší hardware.

## Problémy

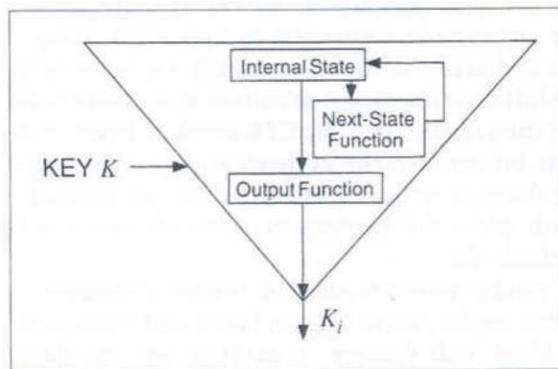
- Nezajišťují samy o sobě integritu.
- Na rozdíl od blokových šifer jsou náchylnější ke kryptoanalytickým útokům, pokud jsou nevhodně implementovány (počáteční stav nesmí být použit opakován) – „problém s inicializačním vektorem“.



Obrázek 13.16: Princip proudových šifer.

### 13.5.1 Rozdelení proudových šifer

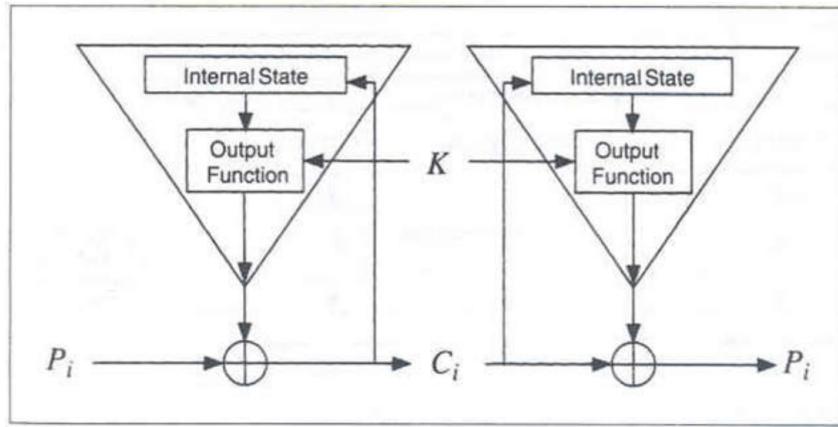
**Synchronní proudové šifry** Proud pseudonáhodných čísel *key stream* je generován nezávisle na vstupním textu nebo zašifrované zprávě. Poté dochází ke kombinaci vygenerovaných čísel se vstupujícím textem (k zakódování) nebo se šifrovaným textem (k dekódování). Nejběžnější formou kombinace keystreamu a vstupního textu je použití operace XOR. Např.: Vernamova šifra, DES v režimu OFB. Pokud se průběhu dešifrování něco ztratí, je konec.



Obrázek 13.17: Princip synchronní proudové šifry.

**Samosynchronizující proudové šifry** Proud pseudonáhodných čísel *key stream* závisí na pevném počtu předcházejících bytů šifrovaného (nebo otevřeného) textu. To znamená, že se šifra dokáže po chybě sama *zotavit* (resynchronize)<sup>1</sup>. Např.: Vigenere Autokey, DES v režimu CFB.

<sup>1</sup>Dnes se nesnážíme dešifrovat poškozená data, pokud nastane chyba v přenosu, vyžádáme si data znova.



Obrázek 13.18: Princip samosynchronizující proudové šifry.

### 13.5.2 Generátory PRNG

Generátory PRNG (*pseudo-random number generator*) generují pseudo-náhodnou posloupnost (*key stream*) z malého klíče (*seed*).

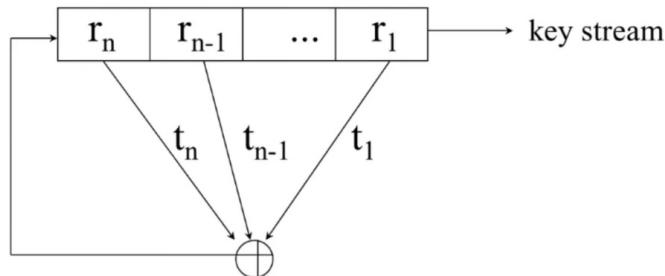
**Blokové šifry v režimu OFB** Blokové šifry v režimu OFB jsou pomalé.

**Linear Feedback Shift Registers (LFSR)** LFSR (posuvný registr s lineární zpětnou vazbou) je posuvný registr, jehož výstup je lineárně závislý na jeho předchozích výstupech a stavu. Mějme

- posuvný registr  $R = (r_1, r_2, \dots, r_n)$ ,
- sekvenci zpětných vazeb  $T = (t_1, t_2, \dots, t_n)$ .

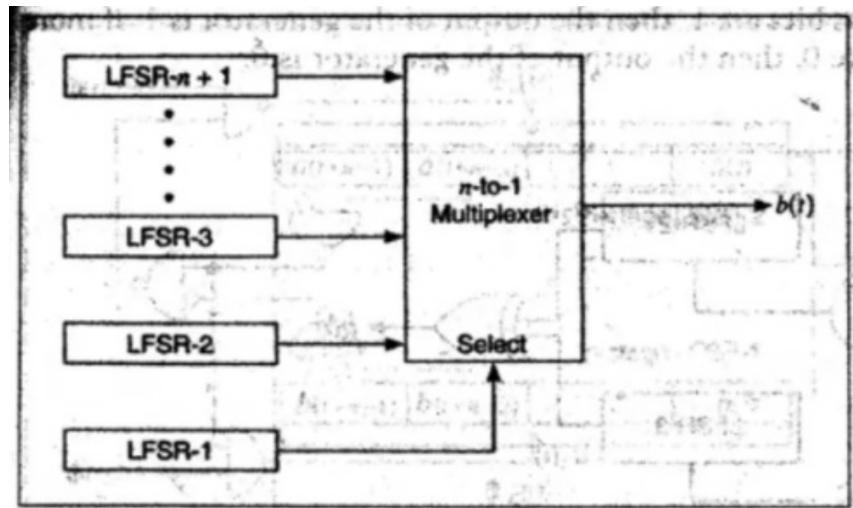
Alternativně lze zapsat polynomem:

$$T(x) = x^n + x^{n-1} + \dots + x + 1$$



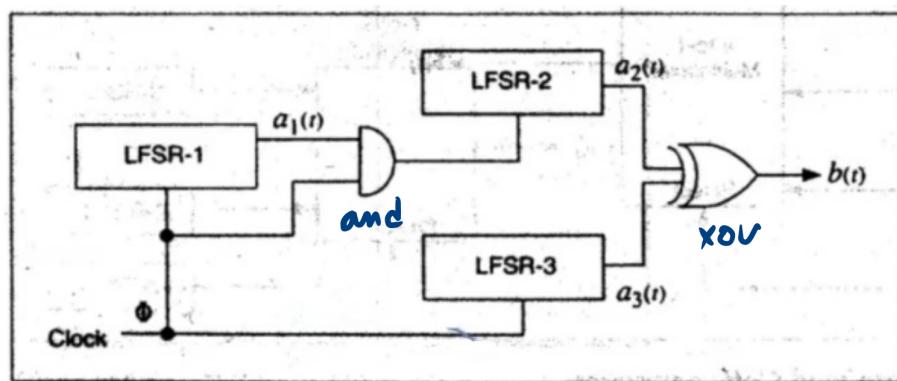
Obrázek 13.19: Příklad LFSR.

**Geffe generátor** Geffe generátor (kombinovaný generátor) je využití 2 a více LFSR propojených multiplexorem.



Obrázek 13.20: Příklad Geffe generátoru.

**Stop and Go generátor** Stop and Go generátor je několik LSFR s různým zdrojem hodin.



Obrázek 13.21: Příklad Stop and Go generátoru.

## Kapitola 14

**KRY – Asymetrická kryptografie,  
vlastnosti, způsoby použití,  
poskytované bezpečnostní funkce,  
elektronický podpis a jeho vlastnosti,  
hybridní kryptografie, algoritmus  
RSA, generování klíčů, šifrování,  
dešifrování.**

### 14.1 Zdroje

- KRY04\_Asym\_MNG.pdf
- KRY\_2021-03-08.mp4
- KRY\_2021-03-22.mp4
- <https://algoritmy.net/article/4033/RSA>

### 14.2 Úvod a kontext

#### Asymetrická kryptografie

- V asymetrické kryptografii se používají páry klíčů (soukromý a veřejný). Soukromý je používán k dešifrování, resp. vytvoření digitálního podpisu. Veřejný je používán k šifrování, resp. ověření digitálního podpisu.
- Každý uživatel generuje svůj pár klíčů. Veřejný klíč je zveřejněn (znají ho všichni), soukromý je držen v tajnosti (zná ho pouze vlastník).
- Všechny asymetrické algoritmy jsou blokové.
- Asymetrické algoritmy jsou pomalejsí než symetrické.

**Způsoby použití** Asymetrická kryptografie lze využít k:

- šifrování,

- digitálnímu podepisování,
- pro výměnu symetrického klíče (*key exchange*).

**Vlastnosti** Vlastnosti symetrické a asymetrické kryptografie<sup>1</sup>.

|                            | Důvěrnost | Autentizace | Integrita | Nepopiratelnost |
|----------------------------|-----------|-------------|-----------|-----------------|
| Symetrická                 | ano       | ?           | ?         | ne              |
| Asymetrická - šifrování    | ano       | ?           | ?         | ne              |
| Asymetrická - podepisování | ne        | ano         | ano       | ano             |
| Asymetrická - kombinace    | ano       | ano         | ano       | ano             |

```

1 # Odesilatel (A):
2 msg = encrypt(msg, SK_A) # nechť msg je zprava k odeslání
3 msg = encrypt(msg, PK_B)
4 send(msg_2)
5
6 # Příjemce (B):
7 msg = receive()
8 msg = decrypt(msg, SK_B)
9 msg = decrypt(msg, PK_A)

```

Výpis 14.1: Kombinace klíčů obou stran u asymetrické kryptografie. Pořadí operací může být i opačné.

**Digitální podpis** Vytvoření digitálního podpisu konkrétních dat pomocí soukromého klíče podepisatele. Každý kdo zná veřejný klíč podepisatele, může pravost podpisu ověřit. Digitální podpis zajišťuje autentizaci, integritu a nepopiratelnost.

**Algoritmy** Algoritmy asymetrické kryptografie se nedají *vymyslet*, musí se objevit. Jsou založeny na těžkých matematických problémech.

- Problém batohu (*knapsack problem*) – MH (Merkle-Hellman)
- Faktorizace čísel – RSA (Rivest-Shamir-Adleman)
- Diskrétní logaritmus – DSA (Digital Signature Algorithm), DH (Diffie-Hellman)
- Eliptické křivky – ECDSA, ECDH

**Problém batohu** Problém batohu je NP-úplný problém kombinatorické optimalizace. Nechť  $x_1, x_2, \dots, x_n$  je množina objektů, každý objekt má svoji cenu  $v_i$  a svoji hmotnost  $w_i$ , dále majme batoh, který má kapacitu  $W$ . Cílem je vybrat takovou množinu objektů, jejichž hmotnost je menší nebo rovna  $W$  a má nejvyšší možnou cenu<sup>2</sup>. Formálně chceme maximalizovat sumu

$$\sum_{i=1}^n v_i \cdot x_i$$

, při splnění

$$\sum_{i=1}^n w_i \cdot x_i \leq W$$

<sup>1</sup>Otazníky – částečně, za předpokladů, ...

<sup>2</sup>Problém má více obdobných variant.

, kde  $x_i \in x_1, x_2, \dots, x_n$ .

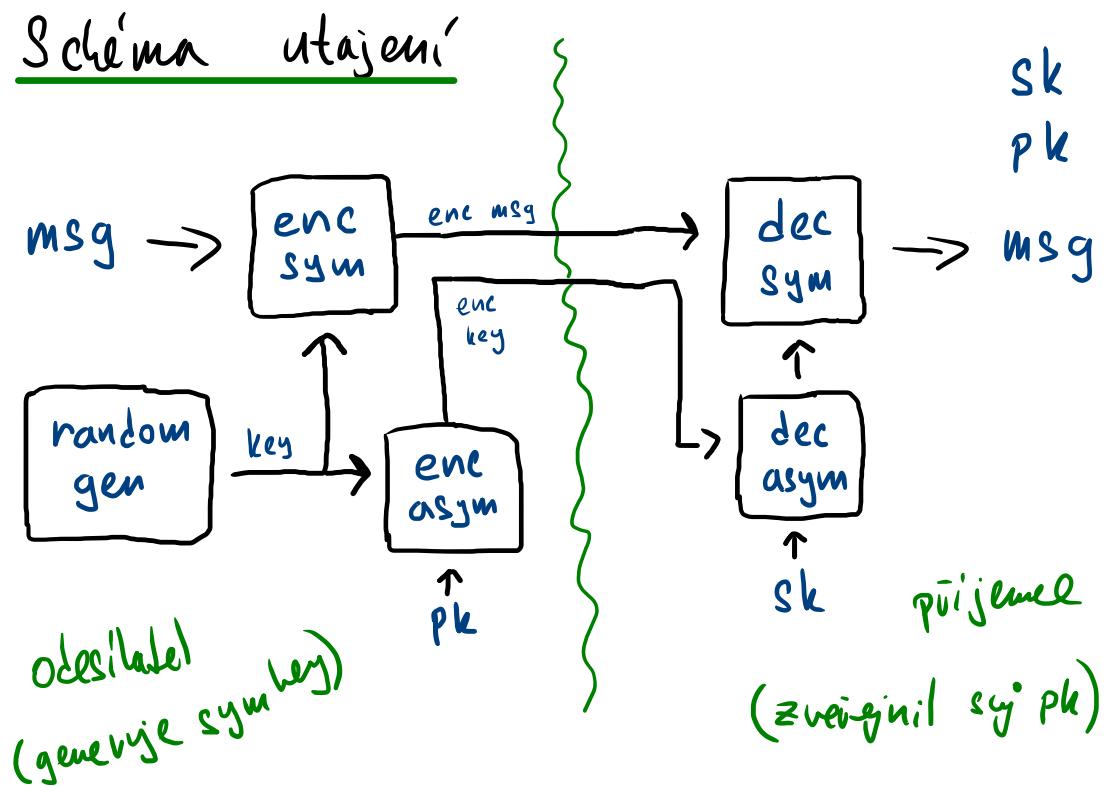
**Faktorizace čísel** Faktorizace čísel označuje problém rozložení čísla na součin menších čísel, v nejčastější podobě pak rozklad celého čísla na součin prvočísel.

**Diskrétní logaritmus** Necht'  $p, g, k, Y$  jsou přirozená čísla, pro něž platí  $Y \equiv g^k \pmod{p}$ . Potom každé číslo  $k$  odpovídající uvedené rovnici nazveme diskrétní logaritmus o základu  $g$  z  $Y$  vzhledem k modulu  $p$ . Tato definice nedefinuje číslo  $k$  jednoznačně, proto se někdy upravuje tak, že ze všech možných diskrétních logaritmů ve smyslu předchozí definice se vybere ten nejmenší.

**Eliptické křivky** Jedná se o matematický aparát, na kterém aplikujeme různé algoritmy (DSA, DH).

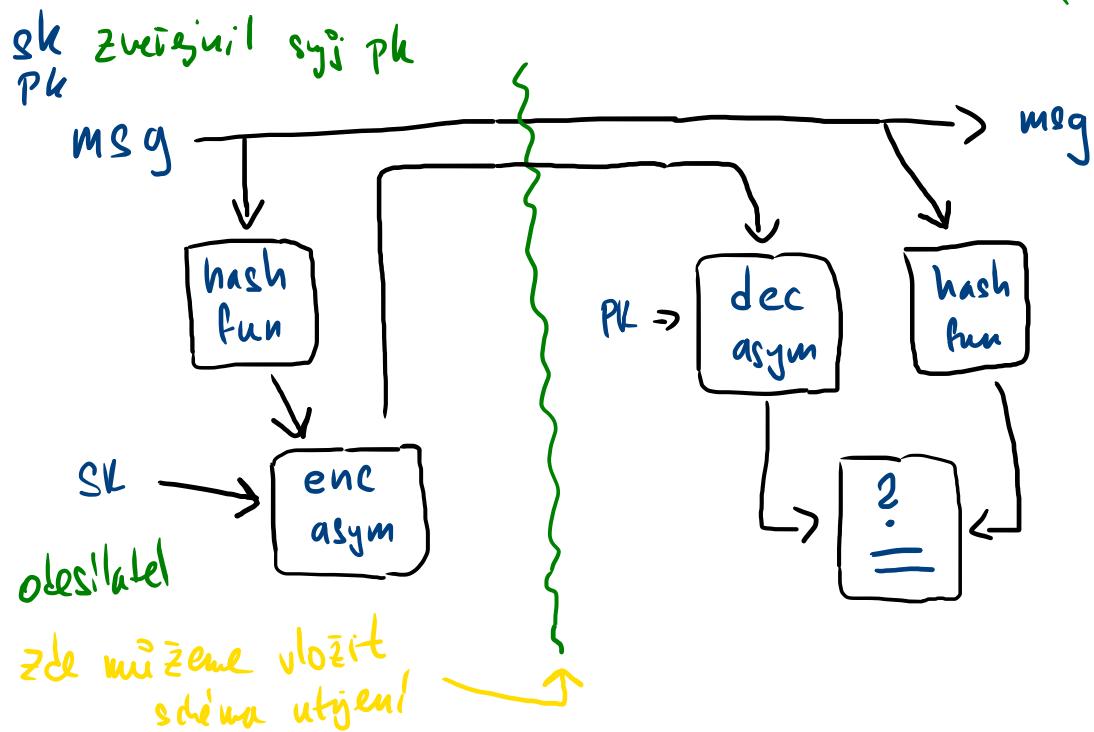
### 14.3 Hybridní kryptografie

Hybridní kryptografie je kombinace symetrické a asymetrické kryptografie, ve které jsou využity přednosti obou (symetrická – rychlá, ale potřeba stejný klíč; asymetrická – pomalá, ale dva klíče). Asymetrická je využita pro bezpečné zaslání symetrického klíče.



Obrázek 14.1: Schéma utajení hybridní kryptografie.

# Schéma digitálního podpisu



Obrázek 14.2: Schéma digitálního podpisu hybridní kryptografie.

## 14.4 RSA

Algoritmus RSA (Rivest-Shamir-Adleman) lze použít jak pro šifrování dat pro digitální podepisování. Je založen na problému faktorizace velkých čísel.

**Klíče** Klíče se skládají z:

- $p, q$  – dvě náhodná soukromá prvočísla,
- $n$  – veřejný modul ( $n = p \cdot q$ ),
- $e$  – veřejný exponent ( $e < \Phi(n)$   $\wedge$   $GCD(\Phi(n), e) = 1$ ), typicky 3 nebo  $2^{16} + 1^3$ ,
- $d$  – soukromý exponent,
- musí platit vztah:  $e \cdot d \bmod \Phi(n) = 1$ .

Veřejný klíč  $PK = (n, e)$ , soukromý klíč  $SK = (n, d)$ .

**Postup generování** Postup generování klíčů:

1. vygenerovat prvočísla  $p$  a  $q$ ,
2. spočítat modul  $n = p \cdot q$ ,
3. spočítat  $\Phi(n) = (p - 1) \cdot (q - 1)$ ,

<sup>3</sup> $GCD$  – největší společný dělitel

4. zvolit veřejný exponent  $e < \Phi(n) \wedge GCD(\Phi(n), e) = 1$ ,
5. spočítat soukromý exponent  $d$  tak, že platí  $e \cdot d \bmod \Phi(n) = 1$ .

**Šifrování a dešifrování** Mějme zprávu  $m$  reprezentovanou jako celé číslo a zašifrovanou zprávu  $c$  reprezentovanou také jako celé číslo. Digitální podpis se vytváří stejným způsobem, pouze se prohodí exponenty.

$$c = m^e \bmod n \quad (14.1)$$

$$m = c^d \bmod n \quad (14.2)$$

**Útoky a slabiny** Pokud útočník rozloží číslo  $n$  na činitele  $p$  a  $q$ , tak může dopočítat soukromý klíč. Pokud útočník uhádně hodnotu  $(p-1) \cdot (q-1)$ , tak může dopočítat soukromý klíč. Šifrování malých čísel je zranitelné, proto se používá „předzpracování“ – zarovnání na  $X$  bitů (2048).

### Příklad

1. RSA  $n = 143$ ,  $p = 11$ ,  $q = 13$ ,  $e = 7$ . Vypočítat  $d$ , napsat VK, PK a zašifrovat číslo 9.

$$p = 11$$

$$q = 13$$

$$n = p q = 143$$

$$e = 7$$

$$\Phi(n) = (p-1)(q-1)$$

$$\Phi(143) = 10 \cdot 12 = 120$$

$$e \cdot d \bmod \Phi(n) = 1$$

$$7d \bmod 120 = 1$$

$$d = 103$$

$$VK = (n, e) = (143, 7)$$

$$SK = (n, d) = (143, 103)$$

$$m = 9$$

$$120x + 7y = 1$$

$$gcd(120, 7)$$

$$\begin{aligned} 120 &= 17(7) + 1 \\ 1 &= 120 - 17(7) \end{aligned}$$

$$\begin{aligned} 1 &= 120 - 17(7) \\ -17(7) &= 1 \\ -17 + 120 &= 103 \end{aligned}$$

$$C = m^e \bmod n = 9^7 \bmod 143 = 4782969 \bmod 143 = \underline{\underline{48}}$$

$$m = C^d \bmod n = 48^{103} \bmod 143 = \underline{\underline{9}}$$

Obrázek 14.3: Příklad RSA.

# Kapitola 15

## KRY – Hašovací funkce, klíčovaný haš a MAC a jejich použití a vlastnosti.

### 15.1 Zdroje

- KRY04\_Asym\_MNG.pdf
- KRY\_2021-03-22.mp4

### 15.2 Úvod a kontext

**Hashovací funkce** Hashovací funkce je funkce (resp. algoritmus) pro převod vstupních dat do (relativně) malého čísla. Výstup hashovací funkce se označuje otisk, *fingerprint*, *digest* či *hash*. Jsou jednosměrné a odolné proti kolizím (viz vlastnosti).

**Obecné vlastnosti** Hashovací funkce by měla:

- Být aplikovatelná na argument o libovolné velikosti.
- Mít výstup konstantní délky.
- Dokázat spočítat výstup rychle.

**Neklíčované hashovací funkce** Hashovací funkce má pouze jeden argument – data. Např. MD2, MD4, MD5, SHS, SHA1, SHA2, SHA3.

$$f(data) \rightarrow hash$$

**Klíčované hashovací funkce** Hashovací funkce má dva argumenty – data a klíč. Také se jim někdy říká MAC (*message authentication code*).

$$f(data, key) \rightarrow hash$$

**Lavinový efekt** Lavinový efekt *Avalanche Effect* je žádoucí vlastností kryptografických algoritmů, typicky blokových šifer a kryptografických hašovacích funkcí, kdy se při nepatrné změně vstupu (například převrácení jednoho bitu) výrazně změní výstup (např. převrátí se polovina výstupních bitů). V případě kvalitních blokových šifer by taková malá změna klíče nebo otevřeného textu měla způsobit drastickou změnu šifrového textu.

### 15.3 Kryptografická odolnost hashovacích funkcí

**Vlastnosti z hlediska odolnosti** Hashovací funkce by z hlediska kryptografické odolnosti měly splňovat:

- *First preimage resistance* – Pro konkrétní  $y$  je výpočetně nezvládnutelné najít takové  $x$ , aby platilo  $h(x) = y$ . Útočník má k dispozici konkrétní hash, a snaží se pro něho nalézt zprávu.
- *Second preimage resistance* – Pro konkrétní  $x$  je výpočetně nezvládnutelné najít takové  $x'$ , aby platilo  $h(x) = h(x')$ . Útočník má k dispozici konkrétní zprávu (nemůže si ji zvolit), ke které se snaží nalézt jinou zprávu, která bude mít stejný hash.
- *Collision resistance* – Je výpočetně nezvládnutelné najít libovolnou dvojici  $x, x'$  takovou, aby platilo  $x \neq x'$  a  $h(x) = h(x')$ . Útočník si může zvolit libovolnou zprávu, ke které se snaží nalézt jinou zprávu, která bude mít stejný hash. Pokud platí *collision resistance*, tak platí i *second preimage resistance*.

**Narozeninový problém** V teorii pravděpodobnosti je narozeninový problém úloha vypočítat minimální početnost skupiny lidí, ve které je alespoň 50% pravděpodobnost nalezení dvojice se stejným datem narození. Narozeninovým paradoxem je pak označována skutečnost, že tento počet (23) je mnohem menší než intuitivní odhad. Výsledek je intuitivnější, když uvážíme, že porovnání narozenin bude provedeno mezi všemi možnými dvojicemi jedinců. Při počtu 23 jedinců je třeba uvažovat  $(23 \cdot 22)/2 = 253$  dvojic, což je více než polovina počtu dnů v roce (182, 5).

Jednodušší je nejprve spočítat jev opačný  $\bar{p}(n)$ , tedy pravděpodobnost, že všech  $n$  narozenin je rozdílných. Pro  $n > 365$  je 1, jinak:

$$\begin{aligned}\bar{p}(n) &= 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right) = \\ &= \frac{365 \cdot 364 \cdots (365-n+1)}{365^n} = \\ &= \frac{365!}{365^n (365-n)!}\end{aligned}\tag{15.1}$$

$$p(n) = 1 - \bar{p}(n)\tag{15.2}$$

**Narozeninový útok** Mějme hashovací funkci, která má  $n$  bitový výstup (celkový počet možných hashů je  $2^n$ ). Útočník vytvoří dokument „přátelská dohoda“ a přibližně  $2^{n/2}$  sémanticky ekvivalentní verzí (úprava bílých znaků, úprava pořadí celků, jiné formulace, ...). Podobně vytvoří dokument „nepřátelská dohoda“ a přibližně  $2^{n/2}$  sémanticky ekvivalentní verzí. S pravděpodobností 0,5 existuje verze „přátelské dohody“ a „nepřátelské dohody“, které mají stejný hash. Pokud takové verze existují, útočník dá oběti podepsat „přátelskou dohodu“  $\Rightarrow$  existuje validní podpis „nepřátelské dohody“.

**Bezpečnostní cíle OWHF** Bezpečnostní cíle OWHF (*one way hash function*). Některé protokoly nevyžadují bezkoliznost, proto má smysl řešit i tento případ.

- Vyžadované vlastnosti: *first preimage resistance* a *second preimage resistance*
- Cíl útočníka: vytvořit *first preimage* nebo *second preimage* (oba úkoly jsou stejně těžké)

- Složitost:  $O(2^n)$  ( $n$  je počet bitů hashe)
- Požadovaná délka:  $n \geq 80$

**Bezpečnostní cíle CRHF** Bezpečnostní cíle CRHF (*collision resistance hash function*).

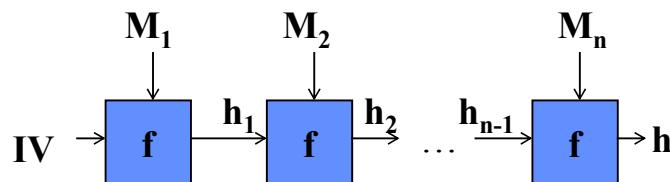
- Vyžadované vlastnosti: *collision resistance*
- Cíl útočníka: vytvořit kolizi
- Složitost:  $O(2^{n/2})$  ( $n$  je počet bitů hashe) (kvůli narozeninovému útoku)
- Požadovaná délka:  $n \geq 160$

**Bezpečnostní cíle MAC** Bezpečnostní cíle MAC (*message authentication code*).

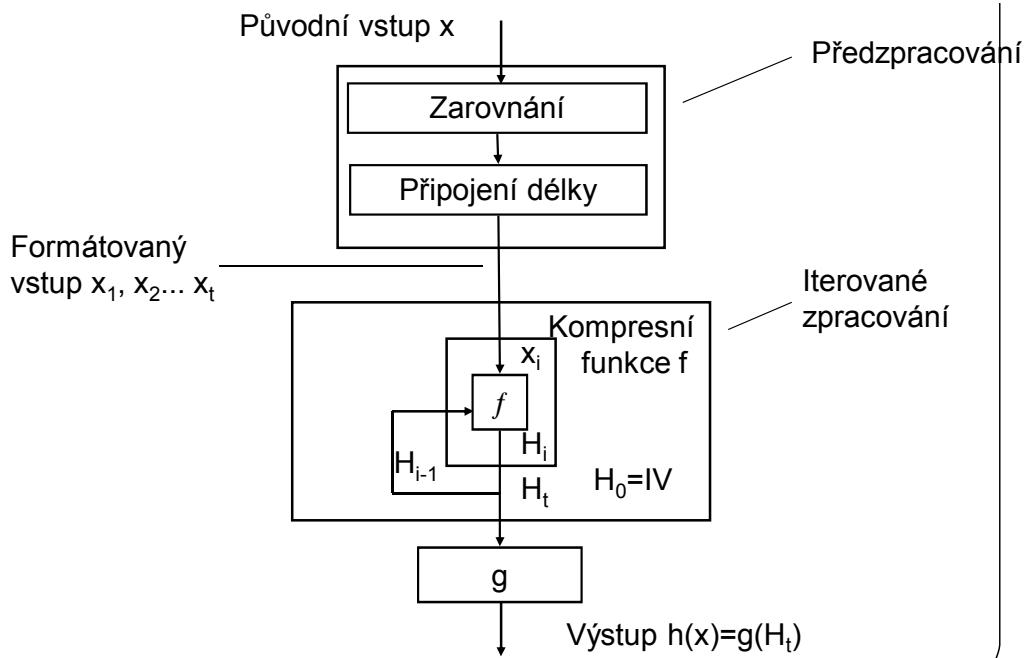
- Vyžadované vlastnosti: *computation resistance, key non-recovery*
- Cíl útočníka (útočník si může vybrat):
  - Vytvořit nový hash, který bude odpovídat nové zprávě
  - Nalézt klíč
- Složitost ( $n$  je počet bitů hashe,  $t$  je počet bitů klíče):
  - Vytvořit nový hash:  $O(\max(2^{-n}, 2^{-t}))$
  - Nalézt klíč:  $O(2^n)$
- Požadovaná délka:  $n \geq 64 \wedge t \geq 64$

## 15.4 Hashovací funkce neklíčované

Nejčastější způsoby sestrojení hashovací funkce neklíčované jsou založené na principu iterace.



Obrázek 15.1: Schéma iterativní neklíčované hashovací funkce. Zpráva je rozdělena na  $n$  částí.  $f$  je tzv. kompresní funkce.  $IV$  je inicializační vektor, resp. konstanta.  $h_1$  až  $h_{n-1}$  jsou mezivýsledky („mezihashe“) a  $h$  je výsledný hash.



Obrázek 15.2: Podrobnější schéma iterativní neklíčované hashovací funkce.

Jednotlivé kroky hashovací funkce:

- Předzpracování – Vstupní data jsou rozdělena na bloky o stejné délce. Je provedeno zarovnání posledního bloku. Je připojena informace o délce zprávy.
- Iterativní zpracování – V iteracích se postupně „přihashovávají“ vstupní bloky. Zpětná vazba pomocí stavové proměnné. Uvnitř kompresní funkce, která z delšího vstupu udělá kratší výstup.
- Postzpracování – Volitelný krok.

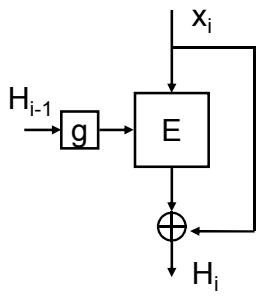
**Merkelova meta-metoda** Necht'  $f$  je kompresní funkce odolná proti kolizím. Hashovací funkce  $h$  na principu iterace využívající kompresní funkci  $f$  je rovněž odolná proti kolizím.

**Merkel-Damgardovo zesílení** Pokud je do vstupu hashovací funkce vložena délka zprávy, tak je zajištěno, že žádná zpráva není prefixem jiné zprávy.

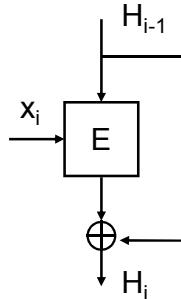
**Zarovnání** Nejednoznačné zarovnání (*ambiguous padding*) – připoj ke zprávě tolik bitů, aby délka zprávy byla násobkem délky bloku. Jednoznačné zarovnání (*unambiguous padding*) – připoj ke zprávě jeden bit a poté proved' nejednoznačné zarovnání.

#### 15.4.1 Hashovací funkce s využitím blokových šifer

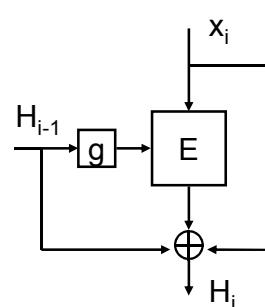
Alternativně lze využít pro konstrukci hashovacích funkcí blokové šifry. Avšak blokové šifry byly navrhovány pro jiný režim činnosti, kterém útočník nezná klíč (a není schopen ho ovlivnit), zná pouze šifrovaný text (ten je schopen ovlivnit). V tomto případě útočník může přímo ovlivňovat hodnoty klíče.



Matyas-Meyer-Oseas



Davies-Meyer

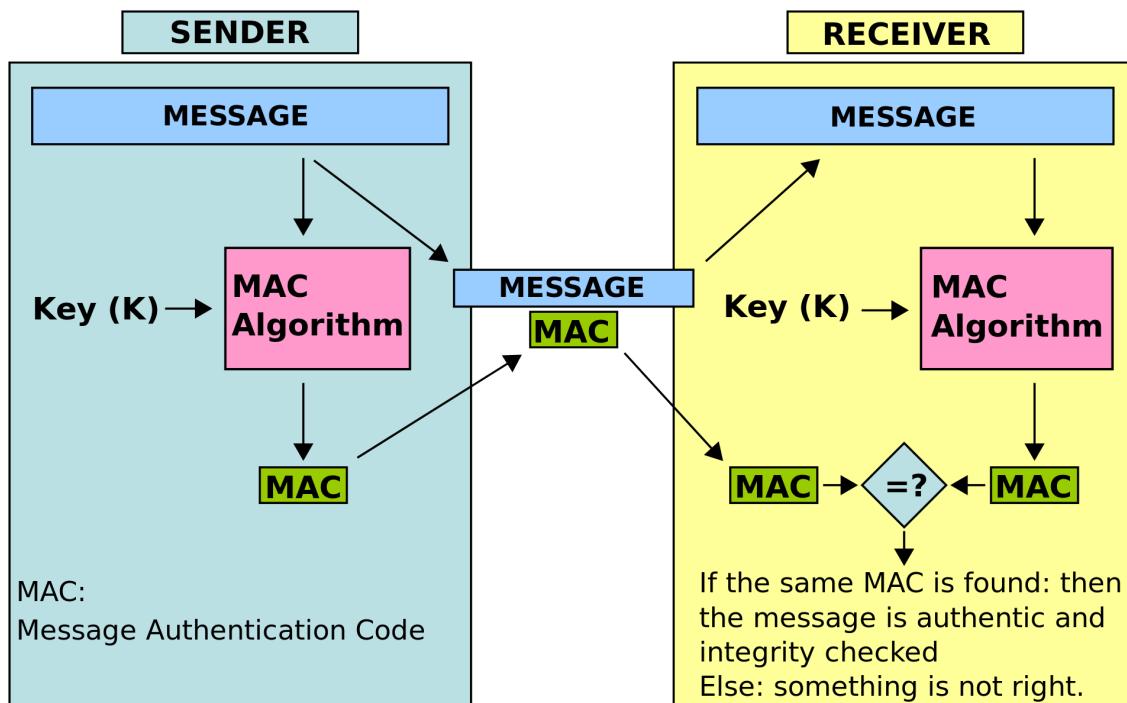


Miyaguchi-Preneel

Obrázek 15.3: Ukázka několika možných způsobů využití blokových šifer pro konstrukci kompresní funkce. S využitím iteračního způsobu lze zobecnit pro celou hashovací funkci.

## 15.5 MAC (*message authentication code*)

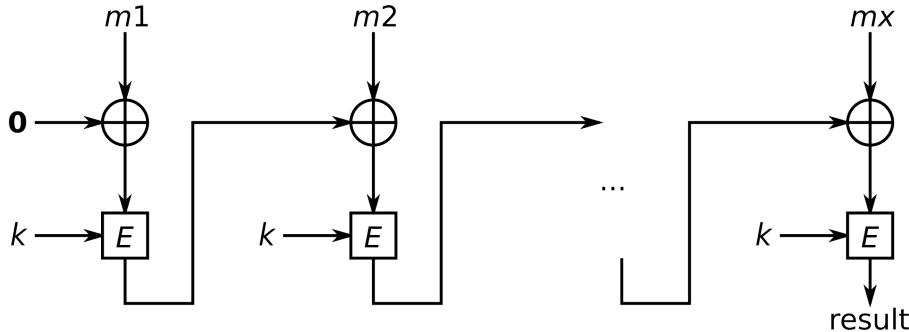
- Rodina hashovacích funkcí  $h_k$ , které jsou parametřitelné klíčem  $k$ .
- Vlastnosti (stejné jako u obecných hashovacích funkcí, pouze rozšířené o klíč):
  - Výstup  $h_k(x)$  lze spočítat rychle, pokud je znám klíč  $k$ .
  - Jsou výpočteně bezpečné – při znalosti dvojice  $(x, h_k(x))$  je výpočteně nemožné spočítat novou dvojici  $(x', h_k(x'))$ , pro  $x \neq x'$ , pokud není znám klíč.
- Využití: zajištění autentizace a integrity (nepopiratelnost zajistit nedokáže).



Obrázek 15.4: Schéma použití MAC. Pokud je stejný MAC výpočítán na straně příjemce, tak má jistotu, že zpráva nebyla po cestě změněna a že zprávu poslal skutečně odesílatel.

### 15.5.1 Sestrojení MAC pomocí blokové šifry v CBC

Pro sestrojení MAC hashovací funkce je využita symetrická bloková šifra v režimu CBC (*cipher block chaining*, šifrová zpětná vazba). Rozdíl oproti CBC šifrování spočívá v tom, že mezivýsledky se zahazují a pracuje se až s posledním blokem. Z něho se vezme určitý počet posledních bitů (podle požadované délky hashe – 32, 48, 64) a ten tvoří výsledný hash (MAC).



Obrázek 15.5: Ukázka sestrojení MAC funkce pomocí symetrické blokové šifry v režimu CBC.

Proč stačí výrazně menší délka? Protože klíč. Útočník sice může vyzkoušet všechny možné hashe, ale bez znalosti klíče, nezjistí, který je ten správný.

### 15.5.2 Sestrojení MAC pomocí neklíčované hashovací funkce

Pro sestrojení MAC hashovací funkce je využita neklíčovaná hashovací funkce. Klíč je připojen ke zprávě a je použita standardní hashovací funkce.

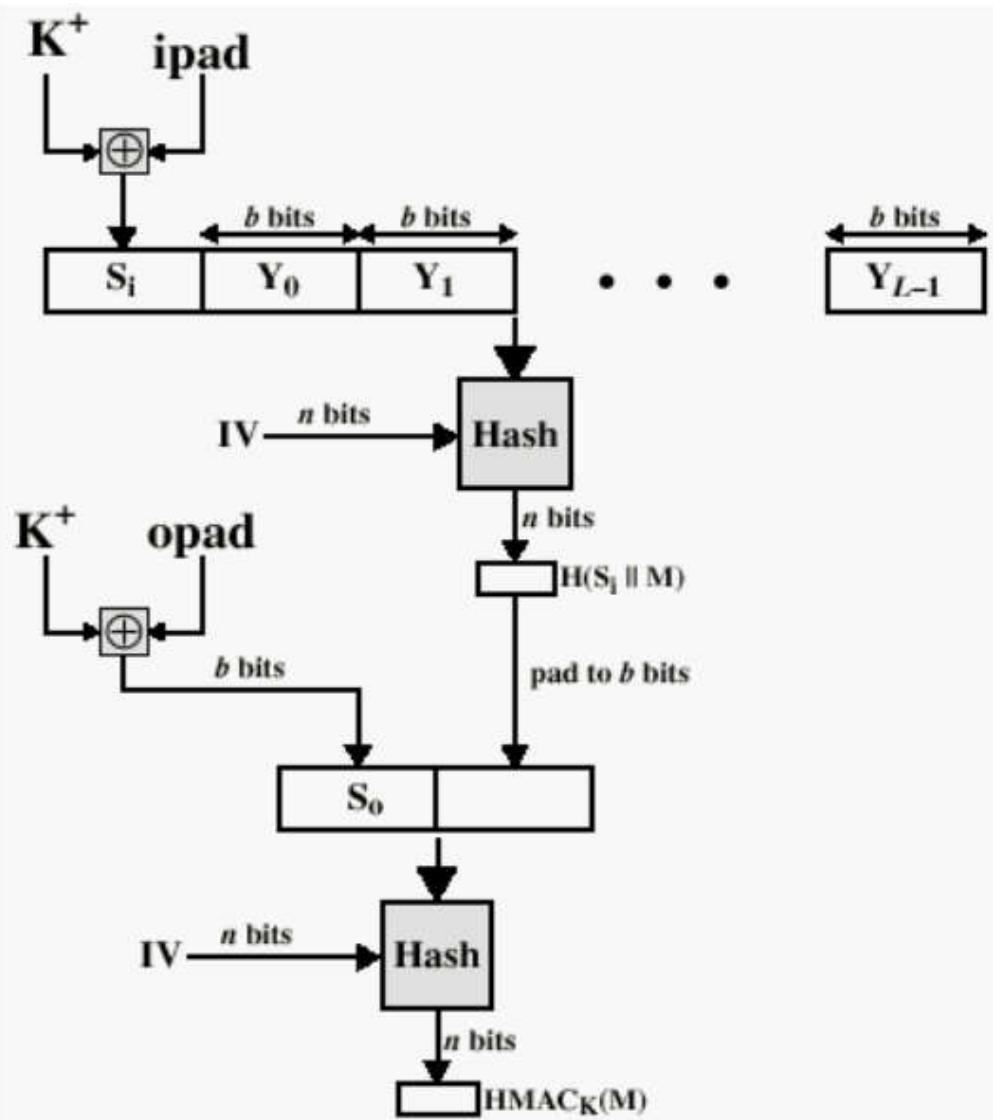
**Secret prefix** Klíč je přidán na začátek zprávy. Formálně:  $H = h(k||x)$ , kde  $H$  je výsledný hash (MAC),  $h$  je hashovací funkce,  $k$  je klíč a  $x$  je zpráva. Útočník může libovolně „při-hashovávat“ další bloky bez znalosti klíče a tím vytvářet nové validní hashe –  $h(k||x||y)$ , kde  $y$  je útočníkova zpráva  $\Rightarrow$  nepřijatelný způsob.

**Secret suffix** Klíč je přidán na konec zprávy. Formálně:  $H = h(x||k)$ . Útočník, který může zvolit  $x$ , může také vytvořit  $x'$ , pro které  $h(x) = h(x')$  se složitostí  $O(2^{n/2})$ , kde  $n$  je délka hashe, bez ohledu na délku klíče  $k$  (narozeninový útok)  $\Rightarrow$  nepřijatelný způsob (útok který je neovlivnitelný délkou klíče).

**Enveloping** Klíč je přidán na začátek i na konec zprávy. Formálně:  $H = h(k||p||x||k)$ , kde  $p$  je zarovnání. Přijatelný způsob. Základ pro algoritmus HMAC.

### 15.5.3 HMAC (*hash function MAC*)

HMAC (*hash function MAC*) je do dnes používaný algoritmus. Specifikuje použití metody enveloping, ale ne, která hashovací funkce se použije.



Obrázek 15.6: Schéma HMAC;  $\text{ipad}$  a  $\text{opad}$  jsou vstupní/výstupní konstanty, které slouží k zarovnání;  $Y_i$  jsou bloky vstupní zprávy;  $IV$  je inicializační vektor.

# Kapitola 16

## KRY – Správa klíčů v asymetrické kryptografii (certifikáty X.509).

### 16.1 Zdroje

- KRY05\_AsymMgmt\_MNG.pdf
- KRY\_2021-03-29.mp4

### 16.2 Úvod a kontext

**Problém se zveřejňováním veřejných klíčů** Jak můžu vědět, že publikovaný veřejný klíč patří opravdu entitě, které patřit má? Je potřeba zajistit autenticitu (pravost) veřejných klíčů – Vytvořit spolehlivou vazbu mezi veřejným klíčem a jménem jeho vlastníka.

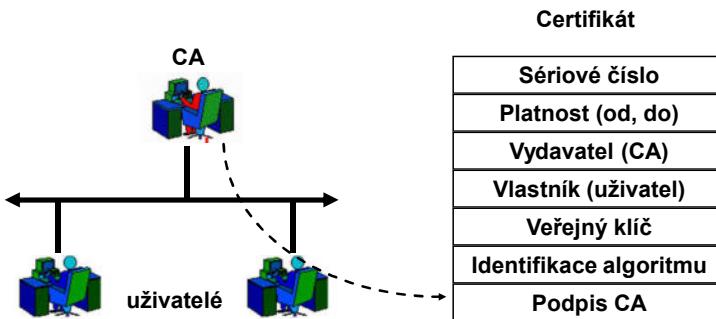
**Systémy založené na veřejném klíči** Systémy založené na veřejném klíči (PKI, *Public Key Infrastructure*) je označení infrastruktury správy a distribuce veřejných klíčů. PKI umožňuje pomocí přenosu důvěry používat cizí veřejné klíče a ověřovat jimi elektronické podpisy bez nutnosti jejich individuální kontroly.

### 16.3 Správa klíčů v asymetrické kryptografii

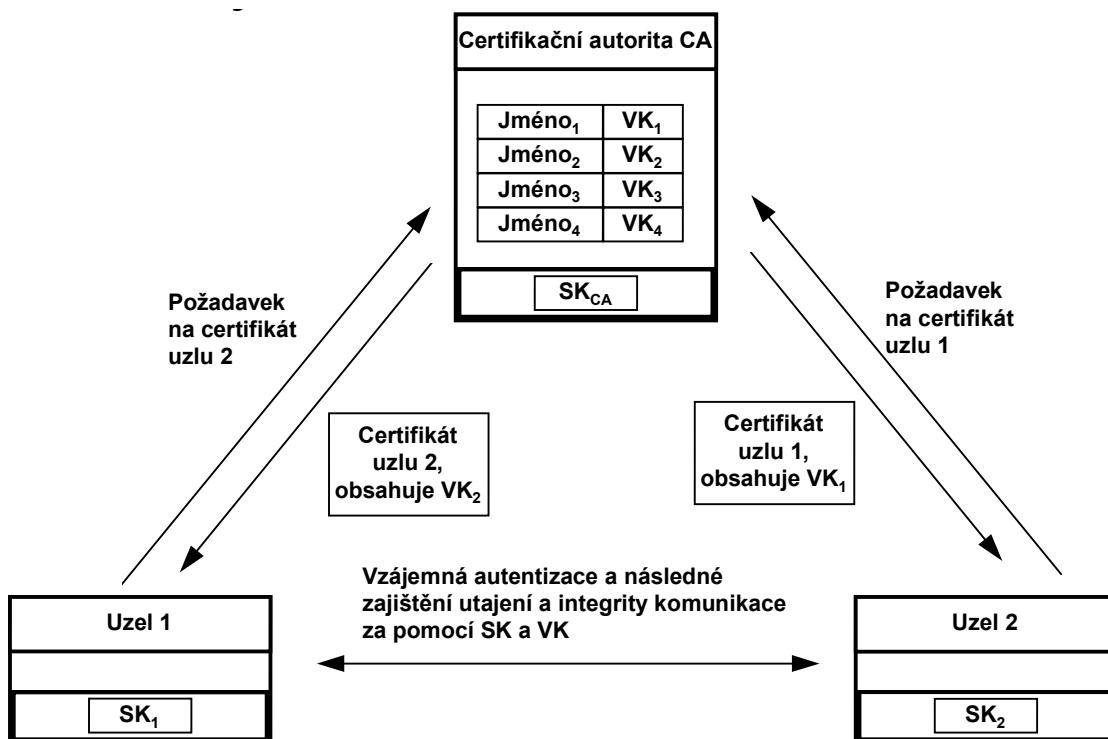
**Certifikát** Certifikace veřejného klíče. Nějaký prostředník (certifikační autorita), kterému důvěřujeme, se zaručuje, že konkrétní veřejný klíč, patří dané entitě.

**Certifikační autorita** Certifikační autorita (CA) je prostředník, který distribuuje certifikáty a které všichni důvěřují. CA negeneruje klíče uživatelům, ty si je generují samy.

**Proces certifikace klíče** CA podepíše veřejný klíč uživatele a jeho další údaje (jméno, doba vydání, doba platnosti, ...) svým soukromým klíčem. Tyto podepsané údaje se nazývají certifikát.



Obrázek 16.1: Příklad certifikátu.



Obrázek 16.2: Příklad navázání bezpečné komunikace mezi dvěma entitami, které mají stejnou certifikační autoritu.

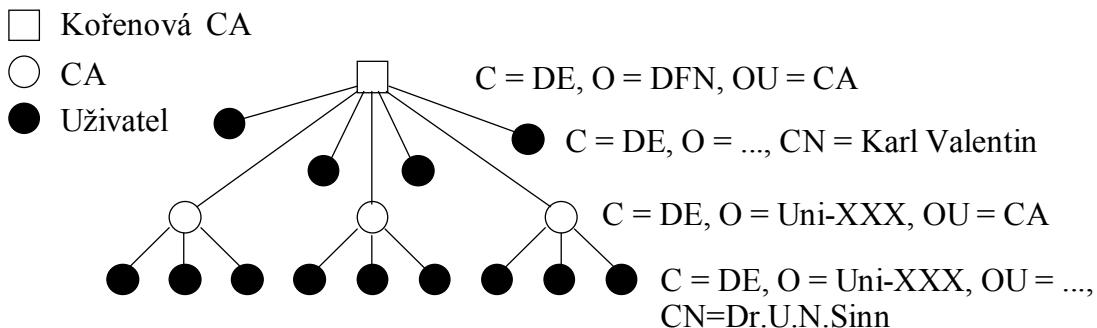
**Navázání bezpečné komunikace** Popis navázání bezpečné komunikace (viz obrázek 16.2):

1. Uzel 1 si vygeneruje soukromý a veřejný klíč.
2. Uzel 1 odešle veřejný klíč certifikační autoritě spolu se svým jménem (a dalšíma informacemi).
3. CA vytvoří certifikát pro uzel 1 – svým soukromým klíčem podepíše veřejný klíč a jméno uzlu 1. CA odešle certifikát uzlu 1. CA odešle svůj veřejný klíč uzlu 1.
4. Pokud uzel 2 chce také odesílat, provede také kroky 1-3.
5. Uzel 1 podepíše soubor a odešle ho uzlu 2 (soubor a podpis).
6. Uzel 2 si musí sehnat certifikát uzlu 1. Existují 3 způsoby jak to udělat.
  - Odesílatel zašle svůj certifikát společně se zprávou.

- Příjemce si vyžádá certifikát odesílatele od certifikační autority.
  - Příjemce si vyžádá certifikát odesílatele od jiné služby (adresářové služby, LDAP).
7. Uzel 2 ověří podpis u certifikátu uzlu 1 veřejným klíčem certifikační autority.
  8. Uzel 2 ověří podpis souboru pomocí veřejného klíče odesílatele (který je v certifikátu).

**Strom certifikačních autorit** Model s jednou globální CA je nemožný (příliš mnoho uživatelů, příliš velké vzdálosti, ...). Proto se používá strom certifikačních autorit. Veřejný klíč CA je certifikován jinou CA. CA nejvýše ve stromu se nazývá **kořenová certifikační autorita**.

- Certifikační autorita má svůj vlastní certifikát, který je podepsaný její certifikační autoritou.
- Koncový uživatel důvěřuje stále pouze jedné entitě – kořenové certifikační autoritě, ale přibývá jedna úroveň ověřování navíc.
- Příjemce dostane zprávu s podpisem. Musí znát certifikát odesílatele (podepsaný CA), certifikát certifikační autority (podepsaný  $CA_{root}$ ) a veřejný klíč kořenové CA<sup>1</sup>.
- Úrovní certifikačních autorit může být více (nejčastěji 1-2).



Obrázek 16.3: Příklad stromu certifikačních autorit. C, O, OU je identifikátor entity.

**Certifikační cesta** Posloupnost certifikátů od certifikátu kořenové CA přes certifikáty dalších CA až k certifikátu komunikující protistrany.

**Zneplatnění certifikátu** Jak zrušit platnost certifikátu? Normálně se zruší sám, až skončí jeho platnost. Pokud je potřeba certifikát zneplatnit před jeho vypršením je třeba využít tzv. revokační seznam (CRL, *certificate revocation list*). Důvody zneplatnění certifikátu:

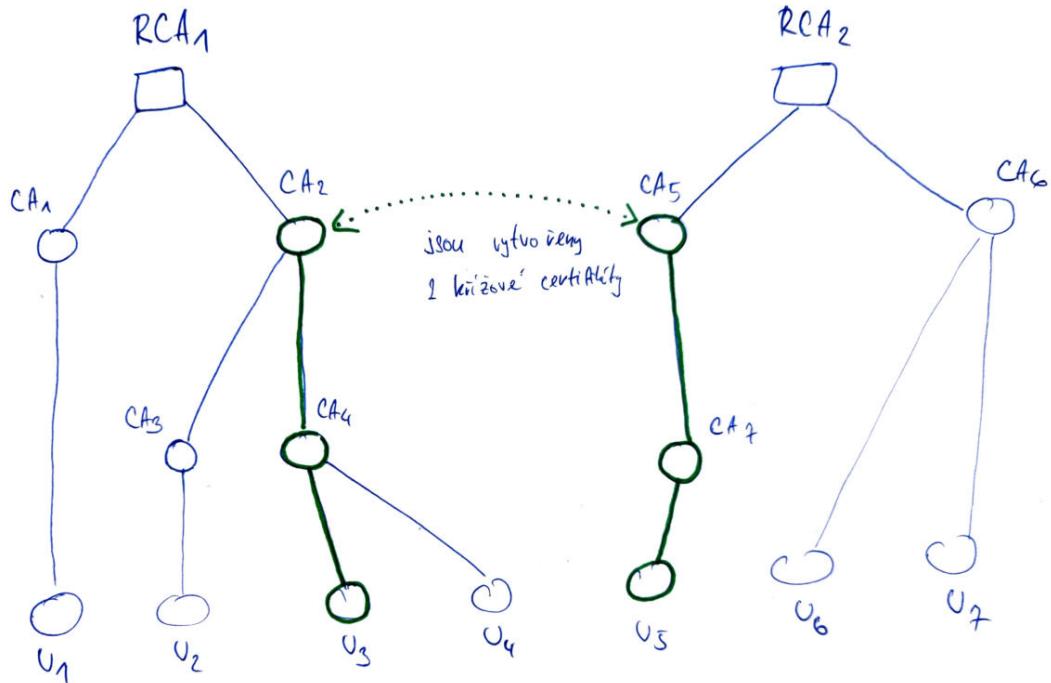
- soukromý klíč uživatele byl kompromitován,
- uživatel ztratil práva, která z certifikátu vyplývají (např. změna zaměstnavatele),
- soukromý klíč CA byl kompromitován (nikdy se nestalo).

**CRL** CRL (*certificate revocation list*) je seznam zneplatněných certifikátů, takových, kterým ještě nevypršela platnost, ale je třeba je zneplatnit. CRL je podepsán CA, které ho spravuje a periodicky aktualizuje (může se zkracovat i růst). Jak se distribuuje:

<sup>1</sup>Veřejný klíč kořenové certifikační autority se z praktických distribuuje ve formě „fiktivního certifikátu“

- *Pull model* – Příjemce certifikátu si dle potřeby stáhne CRL od CA.
- *Push model* – CA pravidelně posílá CRL příjemcům certifikátu.

**Křížový certifikát** V případě, že spolu komunikují uživatelé, kteří nemají společnou kořenovou CA (jejich CA si nedívají), je třeba využít tzv. křížový certifikát. Tím se komplikuje sestavování certifikační cesty, protože je třeba zvážit všechny křížové certifikáty, které jsou k dispozici.



Obrázek 16.4: Příklad křížového certifikátu. Uživatel  $U_3$  chce navázat bezpečné spojení s uživatelem  $U_5$ . Nemají společnou kořenovou CA, proto je třeba využít křížové certifikáty. CA2 vytvoří křížový certifikát pro  $CA_5$  a  $CA_5$  vytvoří křížový certifikát pro  $CA_2$ . Příklad:  $U_3$  pošle podepsanou zprávu  $U_5$ , jak bude vypadat certifikační cesta?

$$U_3 \leftarrow CA_4 \leftarrow CA_2 \leftarrow CA_5 \leftarrow RCA_2$$

## 16.4 Standard X.509

X.509 je standard pro systémy založené na veřejném klíči (PKI). Specifikuje formát certifikátů, formát CRL, parametry certifikátů, metody kontroly platnosti certifikátů, ...

```

1 Certificate ::= SIGNED SEQUENCE {
2     version [0] Version DEFAULT v1988,
3     serialNumber CertificateSerialNumber,
4     signature
5     AlgorithmIdentifier,
6     issuer
7     Name,
8     validity
9     Validity,
10    subject
11    Name,
12    subjectPublicKeyInfo SubjectPublicKeyInfo
13 }
14
15 Version ::= INTEGER {v1988(0) }
16
17 CertificateSerialNumber ::= INTEGER
18
19 Validity ::= SEQUENCE {notBefore UTCTime, notAfter UTCTime }
20
21 SubjectPublicKeyInfo ::= SEQUENCE {
22     algorithm
23     AlgorithmIdentifier,
24     subjectPublicKey
25     BIT STRING
26 }
27
28 AlgorithmIdentifier ::= SEQUENCE {
29     algorithm
30     OBJECT IDENTIFIER,
31     parameters
32     ANY DEFINED BY algorithm OPTIONAL
33 }

```

Výpis 16.1: Příklad definice certifikátu ve formátu X.509.

**Význam položek** Význam položek v definici certifikátu ve formátu X.509:

- Version – Standardně 0.
- Serial number – Sériové číslo certifikátu, spolu se jménem vydavatele jednoznačně identifikuje certifikát.
- Issuer – Jméno vydávající CA.
- Subject – Jméno vlastníka certifikátu.
- Validity – Doba platnosti certifikátu (`notBefore`, `notAfter`). Podpis je platný pouze pokud je datum podepsání v intervalu platnosti každého z certifikátů z certifikační cesty.
- SubjectPublicKeyInfo – Veřejný klíč vlastníka certifikátu a algoritmus, pro který je určen.
- Signature – Jakým algoritmem je certifikát podepsaný CA.

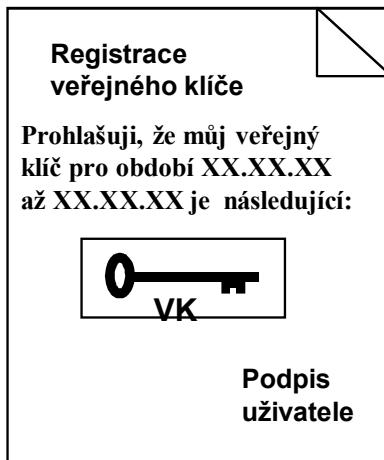
**Prototypový certifikát** Má strukturu certifikátu X.509. Uživatel si vygeneruje tzv. prototypový certifikát, který má standardní strukturu a vyplní informace na nějaké implicitní

hodnoty. Prototyp pošle CA spolu se svým veřejným klíčem, která certifikát dovyplní, podepíše a pošle zpět.

**Registrační autorita** Pokud chce uživatel vydat certifikát, kontaktuje tzv. registrační autoritu (nikoliv přímo CA).

**Míra důvery v certifikát** V praxi chceme více urovní důvěry, než pouze ostrý/žádný (např. chceme vytvořit testovací certifikát). To je řešeno jako rozšíření X.509 přidáním třídy certifikátu (*certification class*). Uživatel chce vydat certifikát od CA jisté třídy.

- Třída 1 – CA vůbec nekontroluje identitu žadatele. Lze jej získat anonymně. Používá se pro testovací certifikáty.
- Třída 2 – Identita žadatele musí být ověřena třetí stranou (notářsky ověřený formulář zasláný poštou).
- Třída 3 – Standardní certifikát. Žadatel musí osobně navštívit CA (resp. registrační autoritu). Osobní ověření totožnosti.
- Třída 4 – Stejně jako 3 a navíc je nutné prokázat oprávnění žadatele požadovat certifikát.



Obrázek 16.5: Příklad žádosti o certifikát.

## Kapitola 17

# PDS – Prerekvizity k ostatním otázkám.

**ISO/OSI model** Referenční model ISO/OSI se používá jako názorný příklad řešení komunikace v počítačových sítích pomocí vrstevnatého modelu, kde jsou jednotlivé vrstvy nezávislé a snadno nahraditelné.

- **Aplikační vrstva (L7, application layer)**

- Zajišťuje zpracování dat na nejvyšší úrovni (reprezentace dat, kódování, řízení dialogu, ...).
- Tvořena procesy a aplikacemi, které komunikují po síti.
- Bývá slučována s prezentační vrstvou (L6, prezentace dat a šifrování) a relační vrstvou (L5, koordinace a komunikace).
- Příklad protokolů:
  - \* Uživatelské – vykonávají služby přímo uživateli (Telnet, SSH, FTP, SMTP, HTTP, ...)
  - \* Systémové – zajišťují síťové funkce (DNS, DHCP, SNMP, BOOTP, ...)

- **Transportní vrstva (L4, transport layer)**

- Rozděluje aplikační data (segmentace) na menší jednotky a zapouzdřuje je do segmentů (TCP) / datagramů (UDP).
- Vytváří logické spojení mezi procesy (přenáší data konkrétní aplikace ze zdrojového zařízení do aplikace na cílovém zařízení).
- Adresace: porty.
- Příklad protokolů: TCP, UDP, DCCP, SCTP, MP-TCP, QUIC

- **Síťová vrstva (L3, network layer)**

- Zapouzdřuje segmenty/datagramy do paketů.
- Řeší směrování.
- Adresace: IP adresa (logická adresa).
- Příklad protokolů: IPv4, IPv6, ARP, RARP, ICMP, IGMP

- **Linková vrstva (L2, data link layer, vrstva síťového rozhraní, network interface layer)**

- Zapouzdřuje pakety do rámců.

- Zajišťuje *hop-by-hop* doručení.
- Adresace: MAC adresa (fyzická adresace).
- Příklad protokolů: Ethernet, Token Ring, FDDI, X.25, Frame Relay

- **Fyzická vrstva** (L1, *physical layer*)

- Zajišťuje přenos bitů přes fyzické médium.

## Adresace

- Port (transportní vrstva, L4)

- Identifikuje aplikaci v rámci zařízení.
- Jak se mění při směrování paketu internetem: zůstává stejný s výjimkou zdrojového portu při překladu NAT.
- Velikost: 16 bit
- Prostor: plochý

- IPv4, IPv6 (sítová vrstva, L3)

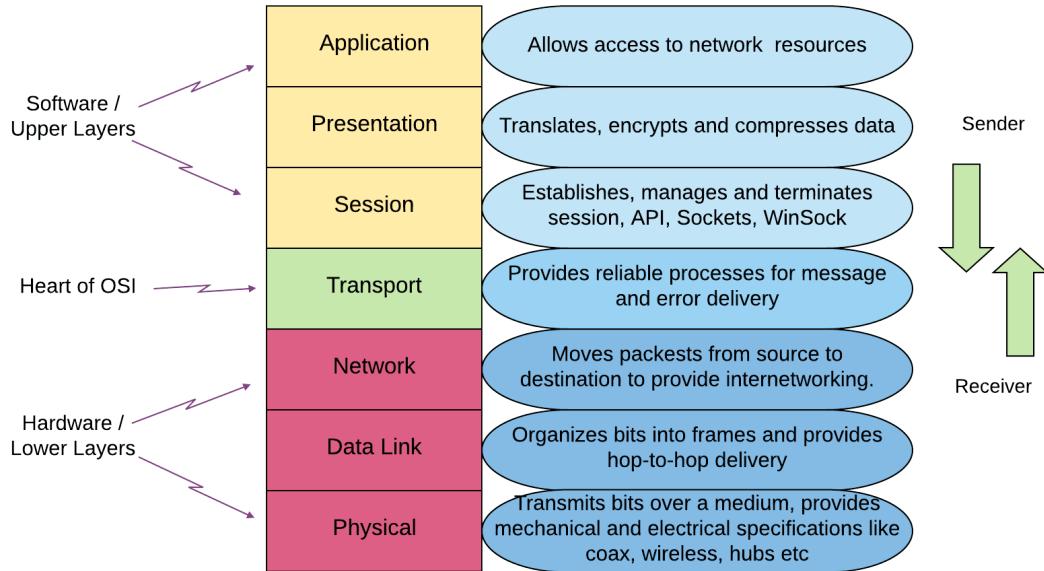
- Identifikuje uzel v rámci sítě (tzv. logická adresace).
- Jak se mění při směrování paketu internetem: zůstává stejná s výjimkou zdrojové IP adresy při překladu NAT.
- Velikost: 32 bit, 128 bit
- Prostor: pseudohierarchie (A, B, C, D, E), pseudohierarchie (prefix + interface ID)

- MAC (linková vrstva, L2)

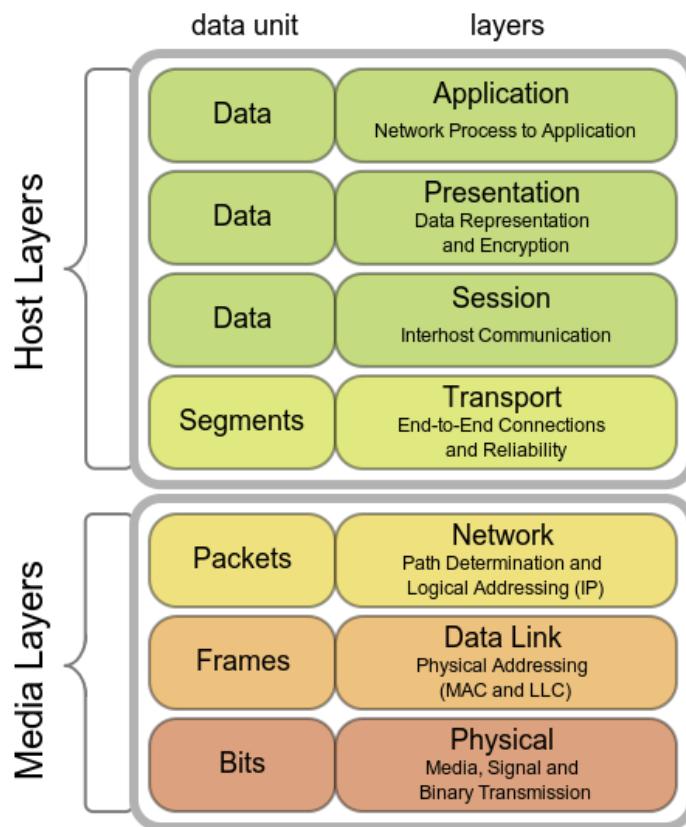
- Identifikuje sítové rozhraní (sítovou kartu).
- Jak se mění při směrování paketu internetem: mění se *hop-by-hop*.
- Velikost: 32 bit, 128 bit
- Prostor: pseudohierarchie (A, B, C, D, E), pseudohierarchie (prefix + interface ID)

## **data, segment, datagram, paket, rámcem, bit**

- Data – aplikační vrstva (L7)
- Segment – transportní vrstva (L4), TCP
- Datagram – transportní vrstva (L4), UDP
- Paket – sítová vrstva (L3)
- Rámcem – linková vrstva (L2)
- Bit – fyzická vrstva (L1)



Obrázek 17.1: Příklad OSI modelu z Linuxhint.



Obrázek 17.2: Příklad OSI modelu z Wiki.

**ACL** ACL (*Access Control List*) je volitelná vrstva zabezpečení, která funguje jako brána firewall pro řízení provozu do jedné nebo více podsítí a z nich.

**NAT** NAT (*Network Address Translation*) je metoda mapování IP adresního prostoru do jiného prostoru (typicky privátní adresy na veřejné adresy). Děje se tak úpravou hlaviček IP paketů během jejich přenosu přes směrovače (úprava zdrojové IP adresy a čísla portu). Směrovač si ukládá čtveřice (WAN\_IP : WAN\_port, LAN\_IP : LAN\_port) aby mohl provádět i překlad zpět.

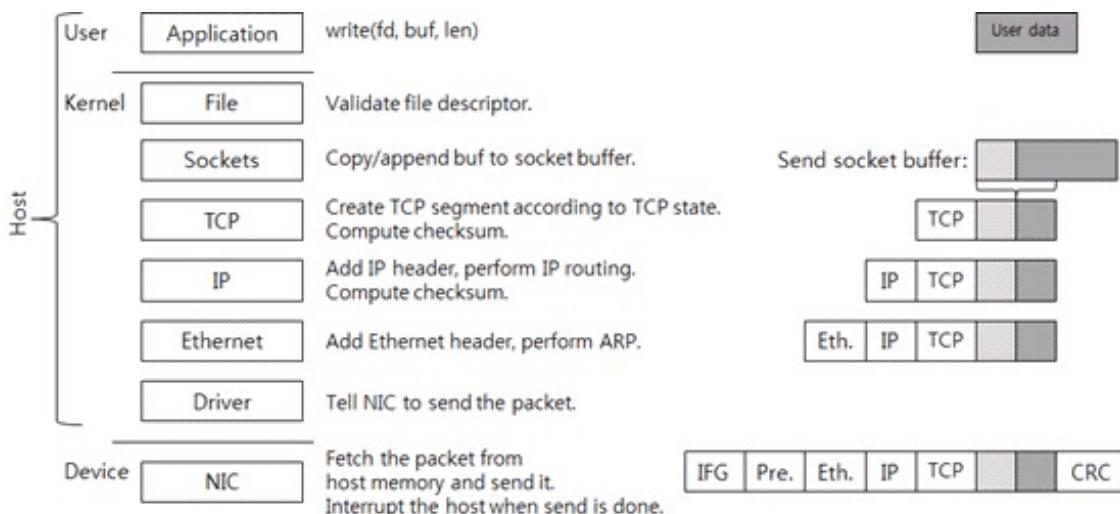
**ARP** ARP (*Address Resolution Protocol*) a RARP (*Reverse ARP*) je protokol, který komunikuje na síťové vrstvě (L3) a zajišťuje „překlad“ IP adres na MAC adresy a obráceně. Pouze pro IPv4, pro IPv6 je pro stejný účel využíván protokol ICMPv6 a zpráva *Neighbor Discovery*. Příklad využití: směrovač potřebuje získat MAC adresu next hopu (zná jeho IP adresu).

**ICMP** ICMP (*Internet Control Message Protocol*) je protokol, který komunikuje na síťové vrstvě (L3) a slouží pro řízení toku a detekce nedosažitelných uzlů.

**MAC adresa** MAC adresa je fyzická adresa zařízení, resp. síťové karty (identifikátor na L2). Na každý síťový port v přepínači/směrovači je jedna síťová karta (a teda i MAC adresa).

**IP adresa** IP adresa je logická adresa zařízení (identifikátor na L3). Pokud má zařízení více síťových karet, má typicky stále pouze jednu IP adresu. Přepínač nemá IP adresu vůbec, pracuje pouze na L2 vrstvě. Směrovač je výjimka a má 2 IP adresy, jednu pro komunikaci v lokální síti (LAN) a druhou pro internet (WAN).

**Síťový tok** Síťový tok je posloupnost paketů (jednosměrná) identifikovaná čtvericí (zdrojová IP, zdrojový port, cílová IP, cílový port).



Obrázek 17.3: Operace na jednotlivých vrstvách ISO modelu.

## Kapitola 18

# PDS – Základní architektury přepínačů, algoritmy pro plánování, řešení blokování, vícestupňové přepínací sítě.

### 18.1 Zdroje

- 04-switching.pdf
- PDS\_2021-03-05.mp4

### 18.2 Úvod a kontext

**CAM tabulka** CAM tabulka je datová struktura v přepínači, která uchovává informace o tom, za jakým fyzickým portem je zařízení s jakou MAC adresou. Jak se plní? Přepínač si ji plní automaticky. V momentě kdy k němu přijde paket, doplní si MAC adresu a port přes který paket přišel. Pokud přepínač neví za jakým portem je cílová MAC adresa, pošle to všem (*flooding*).

**Propustnost** Kolik je možné přenést dat za časovou jednotku.

### 18.3 Obecná architektura přepínače

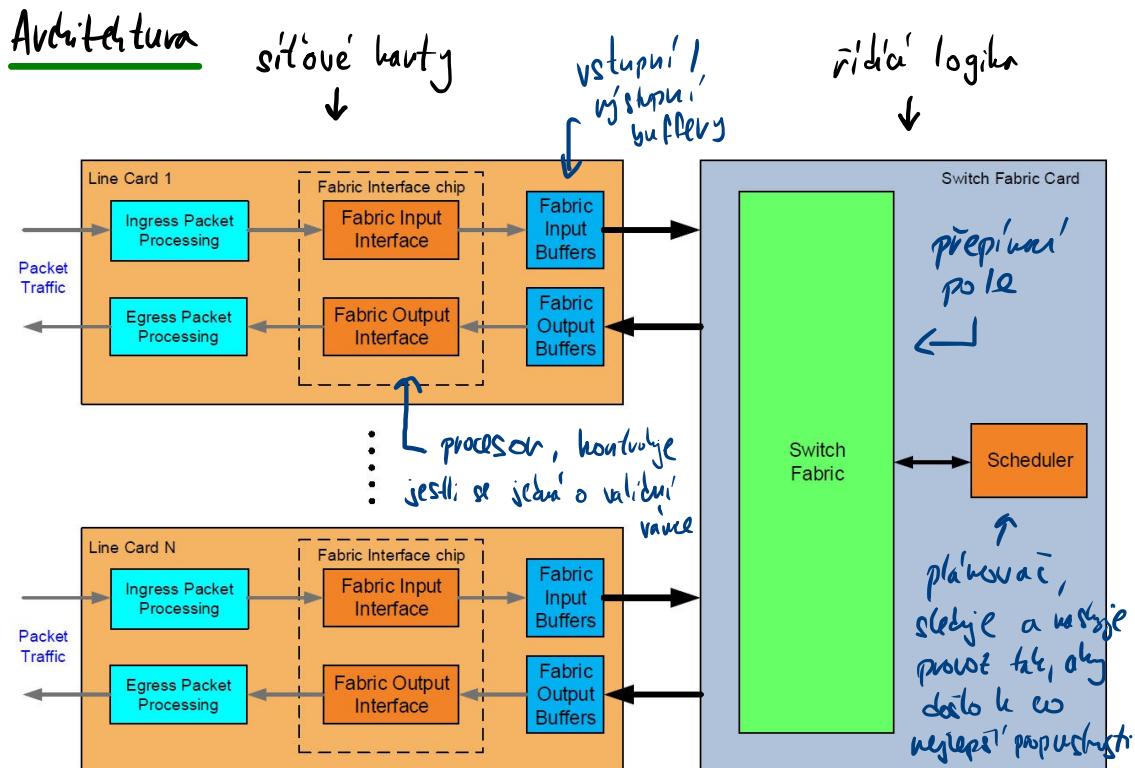
Sít'ový přepínač (*switch*) je aktivní prvek v počítačové síti, který propojuje jednotlivé prvky do hvězdicové topologie. Přepínač obsahuje sít'ové porty (až stovky), na něž se připojují sít'ová zařízení.

- Na jaké vrstvě OSI modelu pracuje? Pracuje s rámci (linková vrstva, L2).
- Na základě čeho provádí přepínání? Na základě cílové MAC adresy.
- Jaké typy přenosů přepíná? Na základě cílové MAC adresy rozslíšuje typ přenosu:
  - broadcast (samé 1),
  - multicast (speciální prefix pro IPv4 a IPv6),
  - unicast (cokoliv jiného).

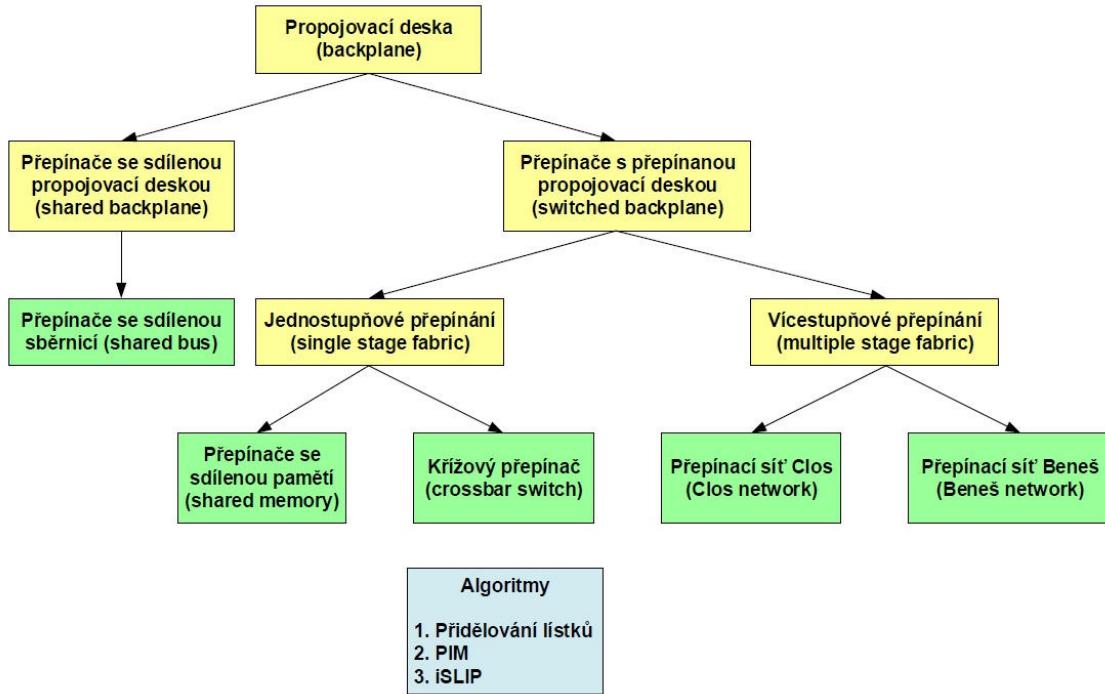
- Co ovlivňuje rychlosť prepínania?
- Hardware (typ pamäti, rychlosť procesoru)
- Logika prepínania
- Šírka pásma cílového rozhraní

### Požadavky na prenos

- Maximálny využití sběrnice (požadavek na plánovač).
- Aby došlo k maximálnemu prenosu dat prepínací logikou (co nejvíce prenosov v rámci taktu – „parallelizace“).
- Spravedlivé pridelenie prenosového pásma (aby boli obsluhované všechny porty).
- Zachovanie pořadí rámcov (ne vždy je možné).



Obrázek 18.1: Obecná architektura prepínače.



Obrázek 18.2: Dělení přepínačů.

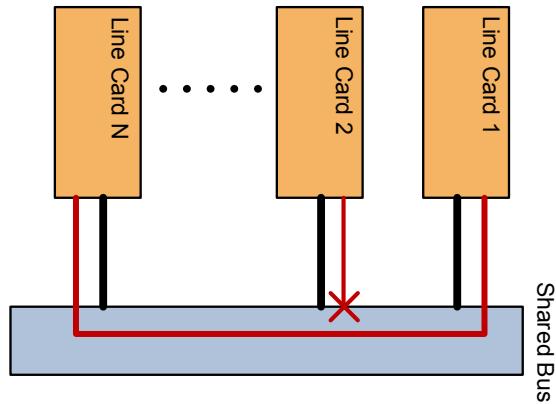
## 18.4 Přepínače se sdílenou propojovací deskou

### 18.4.1 Přepínače se sdílenou sběrnicí

- Může komunikovat pouze jeden port v daný čas, ostatní čekají (řízeno protokolem).
- Lze jednoduše realizovat broadcast a multicast.
- Mějme  $N$  portů (karet), rychlosť každé karty  $R$  bps, taktovací frekvenci  $r$ . Pak rychlosť sběrnice musí být  $N \times R$  a šířka sběrnice

$$W = \frac{R \times N}{r}$$

- Pokud se přidají nové porty nebo se na nich zvýší rychlosť je potřeba zvýšit prospustnost sběrnice („něco za něco“).



Obrázek 18.3: Přepínač se sdílenou sběrnicí.

\* **Úloha č.1:**

- Přepínač má 16 portů o rychlosti 100 Mb/s. Jaká je potřebná propustnost sběrnice a jaká musí být šířka sběrnice, aby zvládl požadovaný přenos? Uvažujte taktovací frekvenci sběrnice 40 MHz.

$$r = 40 \text{ MHz} \quad N = 16 \quad R = 100 \text{ Mb/s}$$

propustnost - kolik dat za časovou jednotku

$$\hookrightarrow N \times R = 1,6 \text{ Gb/s}$$

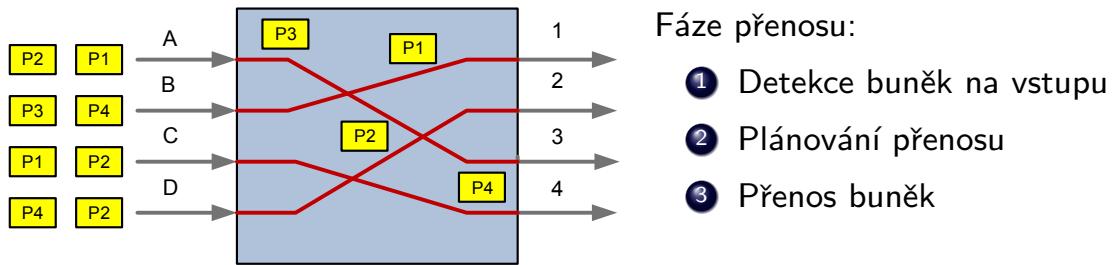
šířka sběrnice - kolik v jednom okamžiku maximálně

$$\hookrightarrow \frac{N \times R}{r} = \frac{16 \cdot 100 \cdot 10^6}{40 \cdot 10^6} = 40 \text{ b}$$

Obrázek 18.4: Přepínač se sdílenou sběrnicí – příklad.

## 18.5 Přepínače s přepínanou propojovací deskou: jednostupňové přepínání

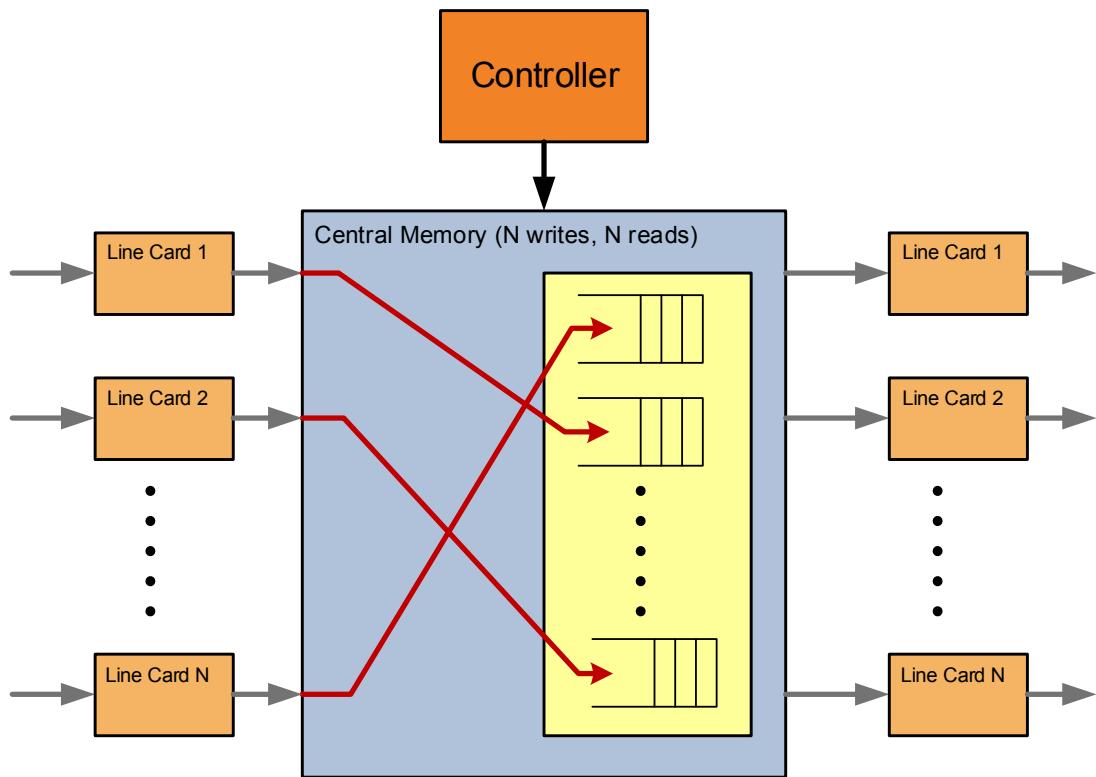
- Umožňuje paralelní přenos rámců.
- Je potřeba plánovač (*scheduler*), který plánuje přenos.
- Provedení přenosu: z bufferů vstupních karet se data přenesou do bufferů výstupních karet.



Obrázek 18.5: Činnost přepínače s přepínanou propojovací deskou.

### 18.5.1 Přepínače se sdílenou pamětí

- Centrální sdílená paměť, která obsahuje frontu pro každý výstupní buffer (každý port).
- Pro jeden přenos se musí 2x přistoupit do paměti (zápis a čtení).
- Zde narážíme na současné technologické limity pamětí (rychlosť přístupu).



Obrázek 18.6: Činnost přepínače s přepínanou propojovací deskou a se sdílenou pamětí. Obsahuje:  $N$  vstupních karet, řadič, sdílenou paměť s  $N$  frontami,  $N$  výstupních karet.

- Necht'  $R$  je rychlosť síťové kart,  $N$  je počet síťových karet, pak celková propustnosť přepínače je  $BW = 2 \times N \times R$  (dva přístupy do paměti).
- Mějme data o velikosti  $C$ , pak doba přenosu je  $t = C \div BW[s]$ .

### Úloha č.2:

- Jaká je potřebná doba přístupu do paměti u přepínače se sdílenou pamětí, který má 32 portů o rychlosti 1 Gb/s a velikost ukládané buňky je 40 bytů? Jak se tato doba změní pro rychlosti portů 10 Gb/s a 40 Gb/s?

$$C = 40 \text{ B} \quad N = 32 \quad R = 1 \text{ Gb/s}$$

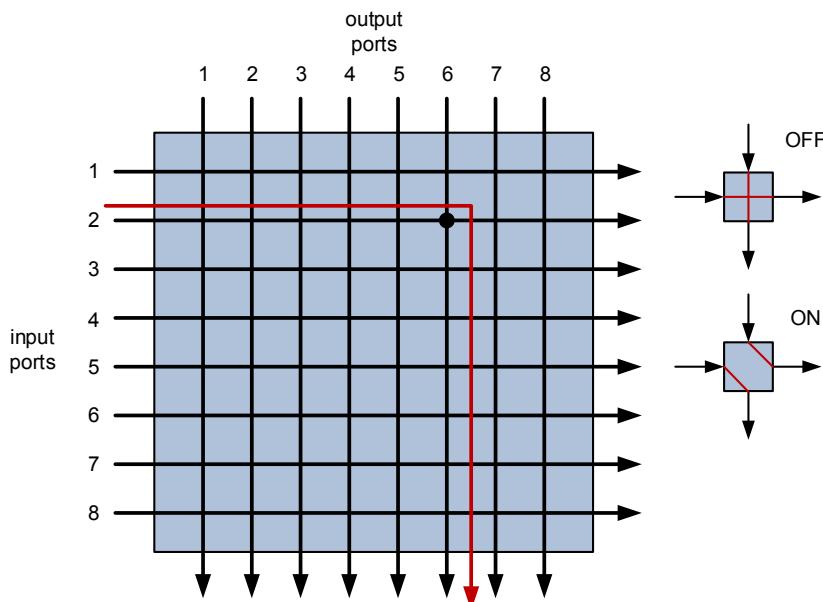
$$t = \frac{40 \cdot 8}{2 \cdot 32 \cdot 1 \cdot 10^9} = 5 \cdot 10^{-9} \text{ s} = 5 \text{ ns}$$

pro 10 → 0,5 ns  
pro 40 → 0,125 ns

Obrázek 18.7: Přepínač s přepínanou sběrnicí a se sdílenou pamětí – příklad.

### 18.5.2 Křížové přepínače (*crossbar*)

- Založený na jednom velkém propojovacím poli.
- Interně neblokující.
- Nativní podpora multicastu.
- Je potřeba  $N^2$  propojení.
- Chybí redundance: pouze jedna cesta mezi vstupem a výstupem.



Obrázek 18.8: Přepínač s přepínanou sběrnicí a křížovými propoji.

## 18.6 Hledání spojů v křížovém přepínači

### 18.6.1 Problém párování

V kontextu přepínání v křížovém přepínači se jedná o problém párování v bipartitním grafu, kde množina vstupních portů je  $V_1$ , množina výstupních portů je  $V_2$  a hrany  $E$  jsou propoje vstupů a výstupů. Množina  $M$  se nazývá párování, pokud  $M \subseteq E$ , kde žádné dvě hrany z  $M$ , nemají společný vrchol.

**Bipartitní graf** Nechť  $G = (V, E)$  je graf. Graf G je bipartitní, pokud platí  $V = V_1 \cup V_2 \wedge V_1 \cap V_2 = \emptyset \wedge \forall e = \{u, v\}, e \in E : u \in V_1 \wedge v \in V_2$ . Platí li navíc  $E = V_1 \times V_2$ , pak je graf úplný bipartitní.

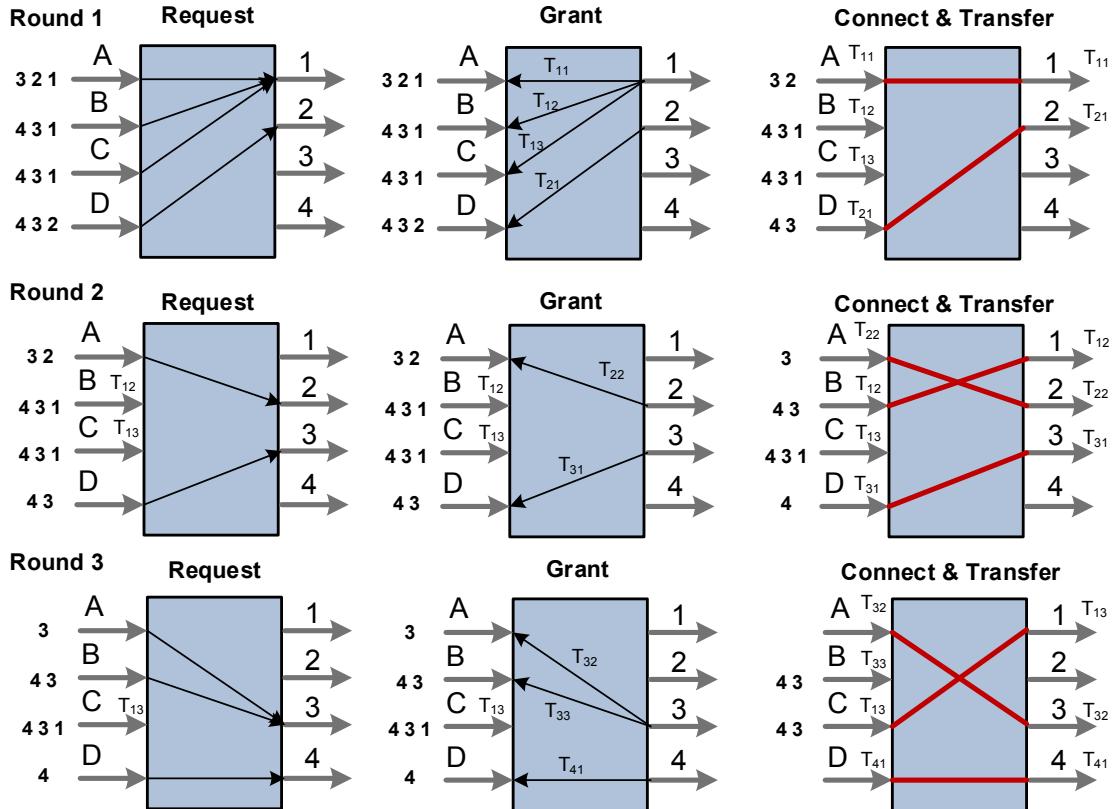
**Největší párování** Největší párování (*maximum matching*) je párování, které má největší počet hran (globální maximum). Složitost:  $O(N^{5/2})$ .

**Maximální párování** Maximální párování (*maximal matching*) je párování, kdy nelze přidat žádnou další hranu (lokální maximum). Složitost:  $O(N + E)$ .

Jelikož algoritmy pro hledání maximálního párování mají výrazně menší časovou složitost, volíme pro hledání spojů v křížovém přepínači ty.

### 18.6.2 Algoritmus přidělování lístků

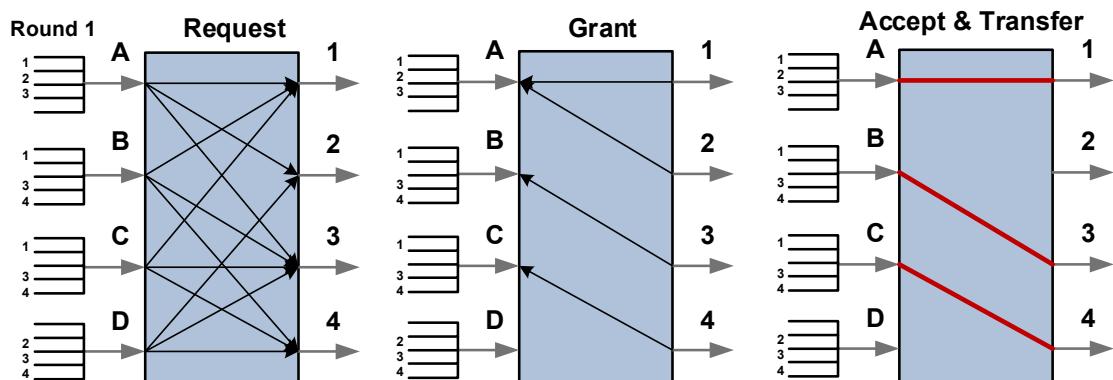
- Princip:
  - Výstupní port  $Q$  obsluhuje frontu požadavků na propojení.
  - Požadavek na vstupu  $P$  dostane od  $Q$  číslo  $T_{QX}$  pro obsloužení portu  $Q$  s pořadím  $X$ .
- Fáze:
  1. Žádost o lístek
  2. Přidelení lístku
  3. Propojení vstupů s výstupy a přenos
- Hodnocení:
  - Umožňuje přenos rámců s proměnlivou délkou
  - **Blokování na začátku fronty** (*head of line blocking*) – Situace, kdy více vstupů chce stejný výstup. Musí se zpracovat ve více kolech. Řešení: Virtuální výstupní fronty (každý vstupní port má frontu pro každý výstupní port). Tím může dojít k předbíhání rámců. Viz další algoritmy.



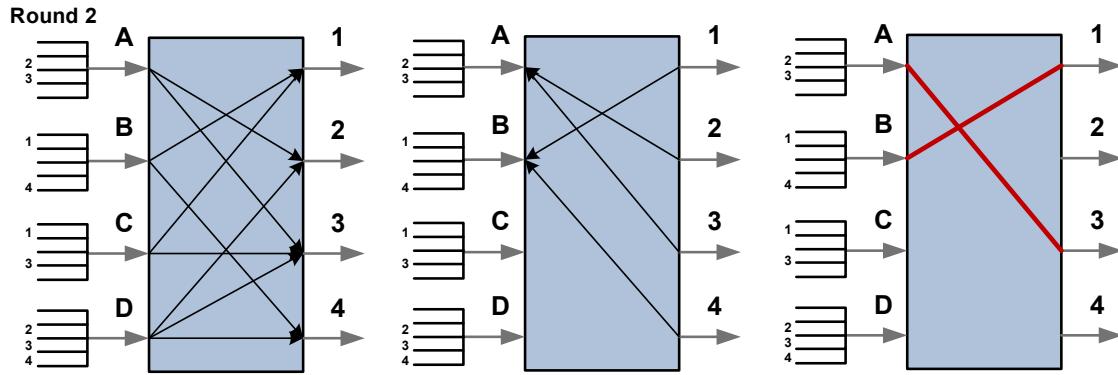
Obrázek 18.9: Příklad činnosti algoritmu přidělování lístků.

### 18.6.3 Algoritmus PIM (*Parallel Iterative Matching*)

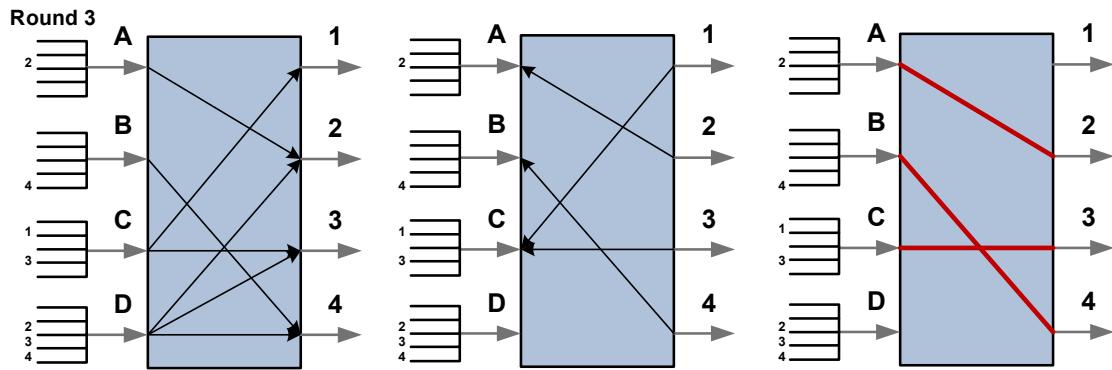
- Algoritmus přidělování lístků + virtuální výstupní fronty + náhodnost.
  - Posílají se požadavky pro všechny pakety ve frontě, nikoliv pouze pro první.
- Hledá maximální párování (pomocí náhodné volby, která zabrání vyhledování).
- Soutěžení o porty
  - Výstupní port – více žádostí naráz, jednu vyberu náhodně.
  - Vstupní port – povolení na více výstupů naráz, jeden vyberu náhodně.



Obrázek 18.10: Příklad činnosti algoritmu PIM, kolo 1.



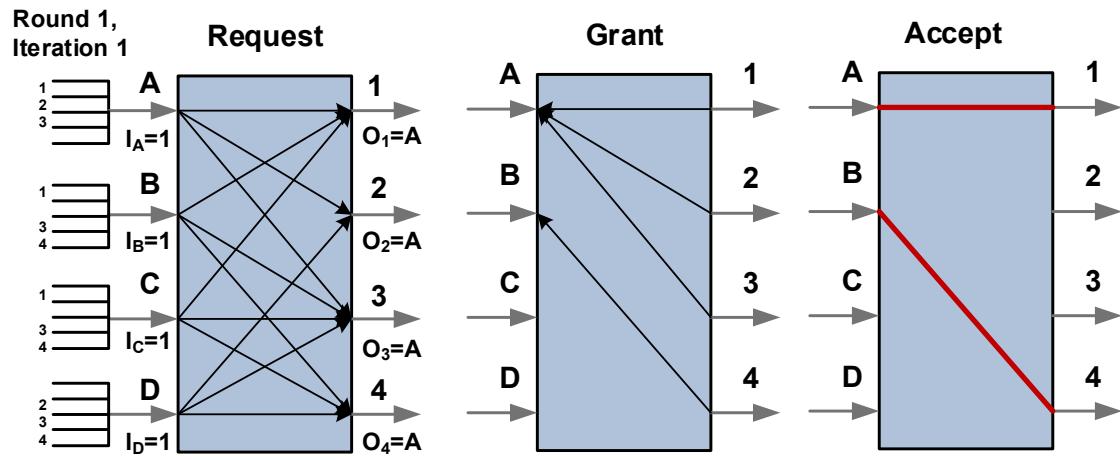
Obrázek 18.11: Příklad činnosti algoritmu PIM, kolo 2.



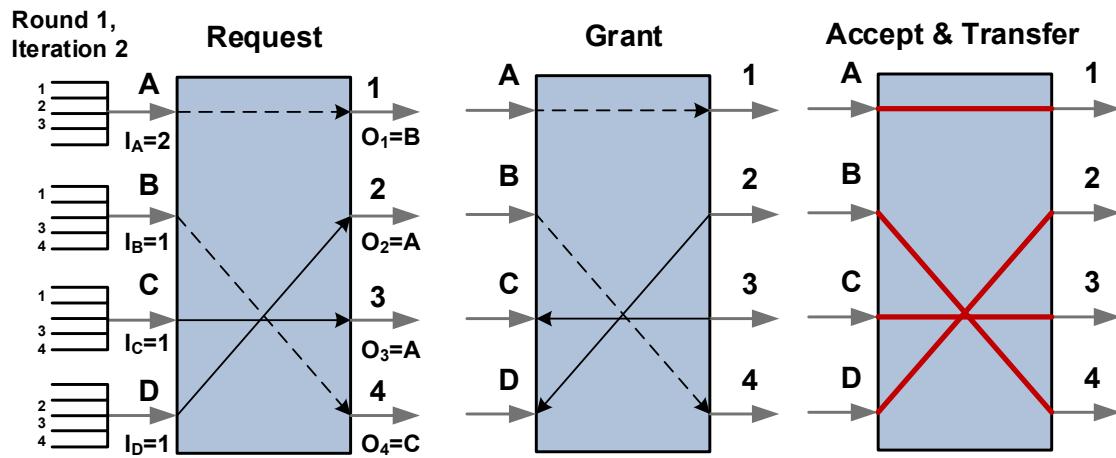
Obrázek 18.12: Příklad činnosti algoritmu PIM, kolo 3.

#### 18.6.4 Algoritmus iSLIP

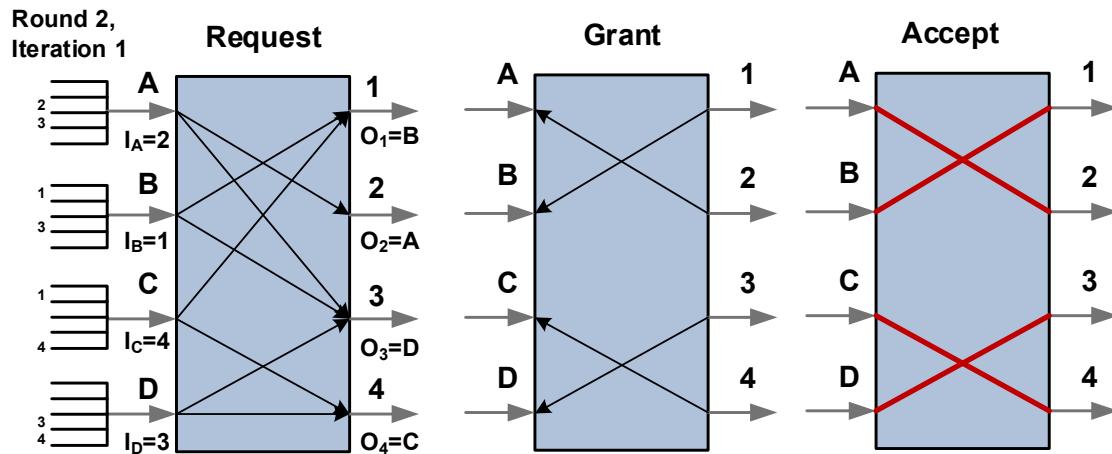
- Algoritmus přidělování lístků + virtuální výstupní fronty + iterace + ukazatele.
- Algoritmus „rotujících ukazatelů“ (vyhneme se náhodnému výběru).
  - Každý vstupní port má ukazatel (počítadlo), podle kterého se určuje priorita výstupu.
  - Každý výstupní port má ukazatel, podle kterého se určuje priorita vstupu.
  - Nová hodnota ukazatele je port na který se posílá / port od kterého se přijímá +1.
- Každé kolo má dvě iterace, resp. 6 fází – *request, grant, accept, request, grant, accept and transfer*.
  - Proč? Protože fáze *transfer* trvá nejdéle (optimalizace).



Obrázek 18.13: Příklad činnosti algoritmu iSLIP, kolo 1.



Obrázek 18.14: Příklad činnosti algoritmu iSLIP, kolo 2.



Obrázek 18.15: Příklad činnosti algoritmu iSLIP, kolo 3.

## 18.7 Přepínače s přepínanou propojovací deskou: vícestupňové přepínání

Vlastnosti jednostupňového křížového přepínače:

- S počtem portů roste kvadraticky velikost přepínacího pole.
- Vnitřní blokování (soupeření o porty).
- Blokování na začátku fronty (*head of line blocking*).
- Součástí přepínání je hledání maximálního párování.

Je možné zvětšit počet portů, aniž by se dramaticky rozšířila velikost přepínacího pole?  
Ano, pomocí vícestupňového přepínání (vstup  $\rightarrow$  vnitřní přepínací obvody  $\rightarrow$  výstup).

- Nemají vnitřní blokování – Nejsou potřeba plánovací algoritmy pro hledání maximálního párování.

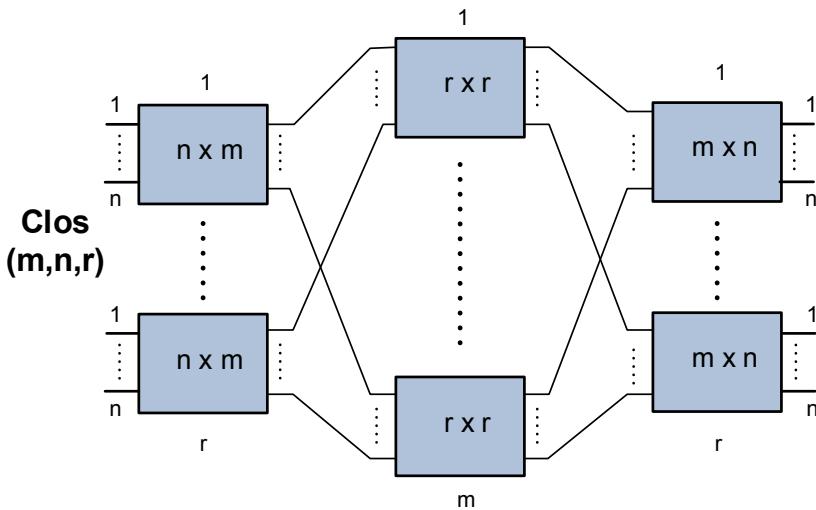
### 18.7.1 Přepínací síť Clos

$Clos(m, n, r)$

- $r$  vstupních bloků s  $n$  vstupy (bloků = křížový přepínač)
- $r$  výstupních bloků s  $n$  vstupy
- $m$  vnitřních bloků s  $r$  vstupy

Closův teorém:

- Pokud  $m \geq 2n - 1$ , pak lze přidat nové propojení vstupu a výstupu bez přeskladání (sít' je neblokující).
  - Pokud všechny požadavky na vstupu jdou na jiné výstupy, pak pro jakoukoliv konfiguraci vstupů a výstupů je síť neblokující (pokud nejdou dva vstupy na jeden stejný výstup).

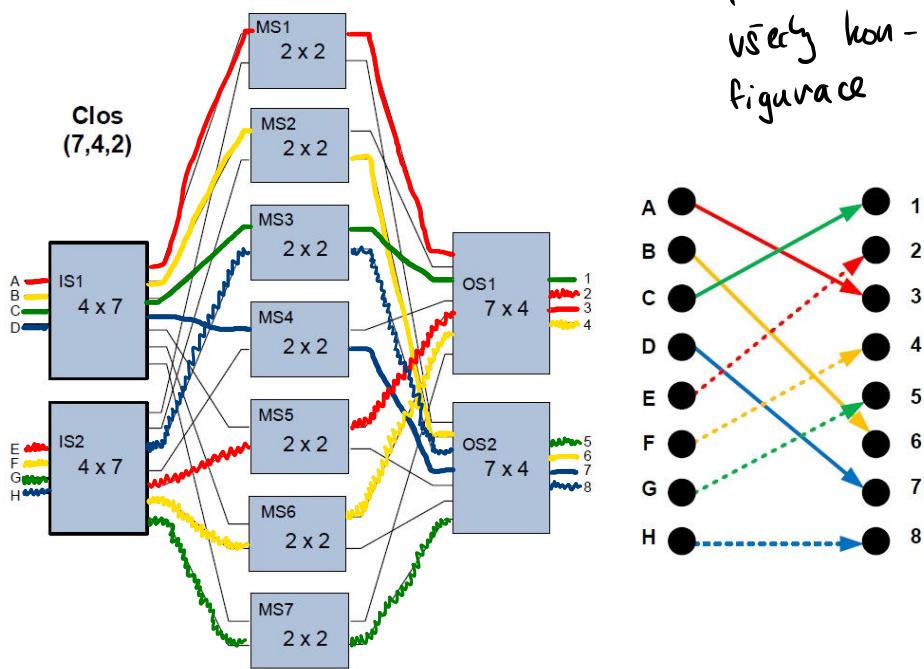


Obrázek 18.16: Obecné schéma Closovy sítě.

## Příklad neblokující sítě Clos(7,4,2): osmiportový přepínač

- Platí Closův teorém, tj.  $m \geq 2n - 1$

• Vždycky najde  
propojení pěs  
všechny kon-  
figurace



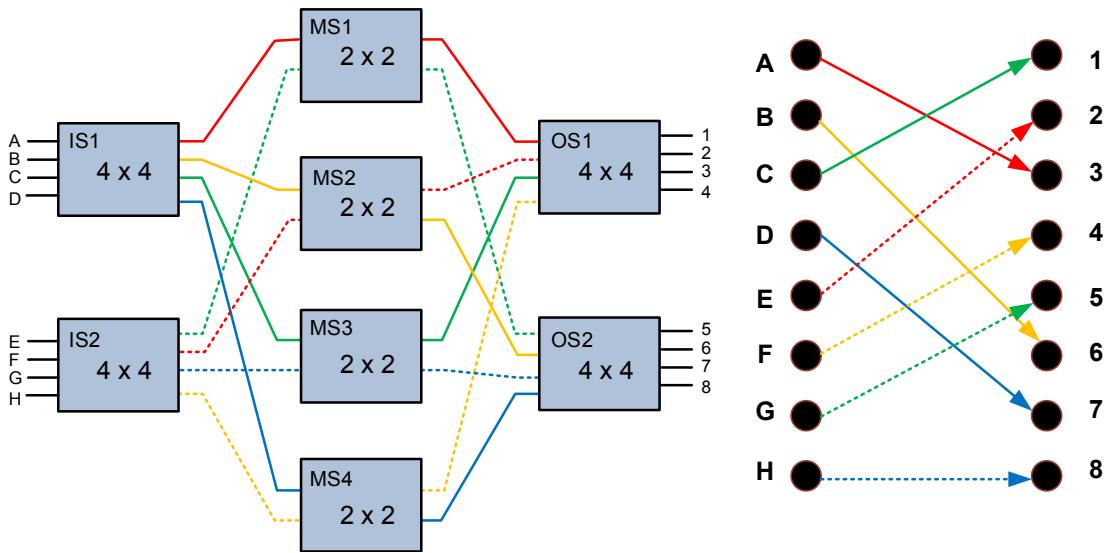
Obrázek 18.17: Příklad Closovy sítě.

### 18.7.2 Modifikovaná přepínací síť Clos

- Pokud  $m \geq n$ , pak je síť neblokující po přeskládání.
- Výpočetní složitost přeskládání je  $O(N \log D)$ , kde  $D$  je počet barev.
- Proč? Při plnění closova teorému je potřeba hodně bloků. Při této podmínce stačí výrazně méně a složitost výpočtu přeskládání je přijatelná.

### Příklad sítě Clos(4,4,2): osmiportový přepínač

- Příklad neblokující konfigurace po přeskládání.

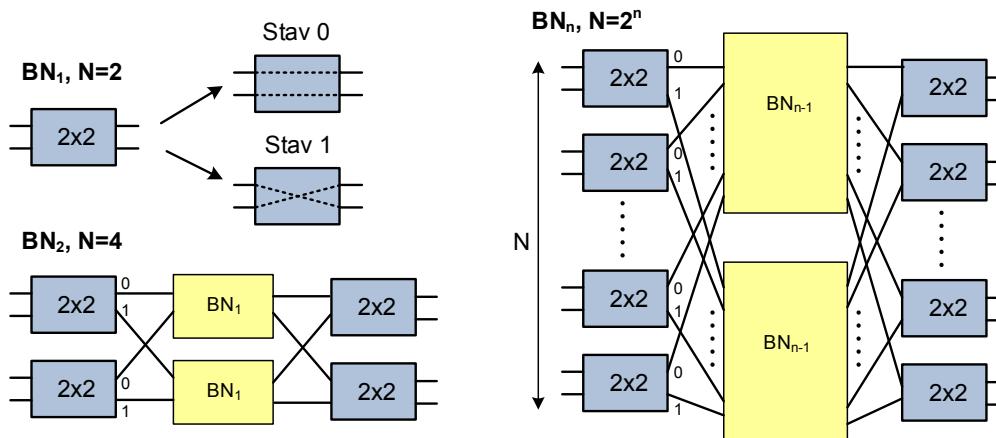


Obrázek 18.18: Příklad modifikované Closovy sítě.

### 18.7.3 Přepínací sít' Beneš

*Benes(n)* také  $BN(n)$

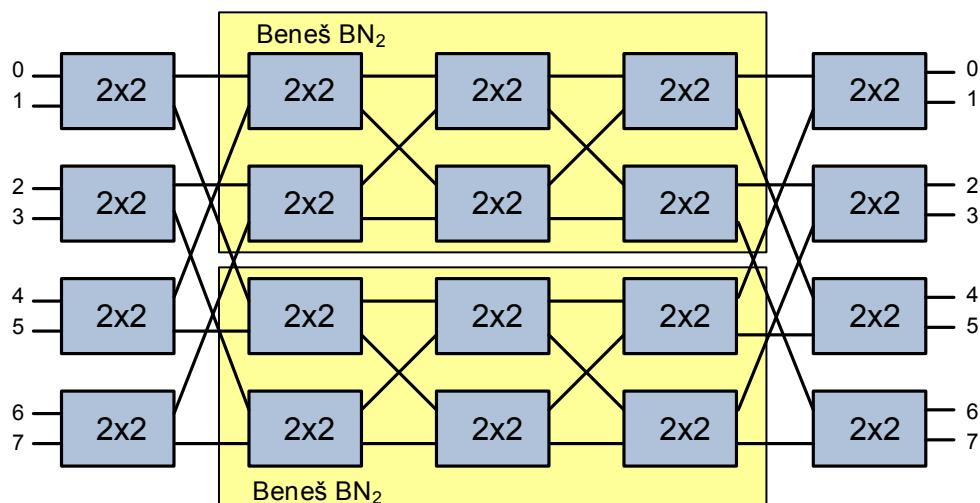
- Základem je modifikovaná síť  $Clos(m = 2, n = 2, r = 1)$ .
- Rekurzivní konstrukce sítě  $Benes(n)$ , kde  $n$  je stupeň rekurze.
- Vstupní a výstupní přepínače velikosti  $2 \times 2$ .
- Celkový počet vstupních/výstupních portů  $N = 2^n$ .
- Prostřední část rekurzivních bloků  $BN(n - 1)$ .
- Neblokující po přeskládání – algoritmus hledání propojení (žádné soupeření o vnitřní uzly nebo výstupní porty).



Obrázek 18.19: Obecné schéma Benešovy sítě.

### Příklad: přepínací síť Beneš BN<sub>3</sub>

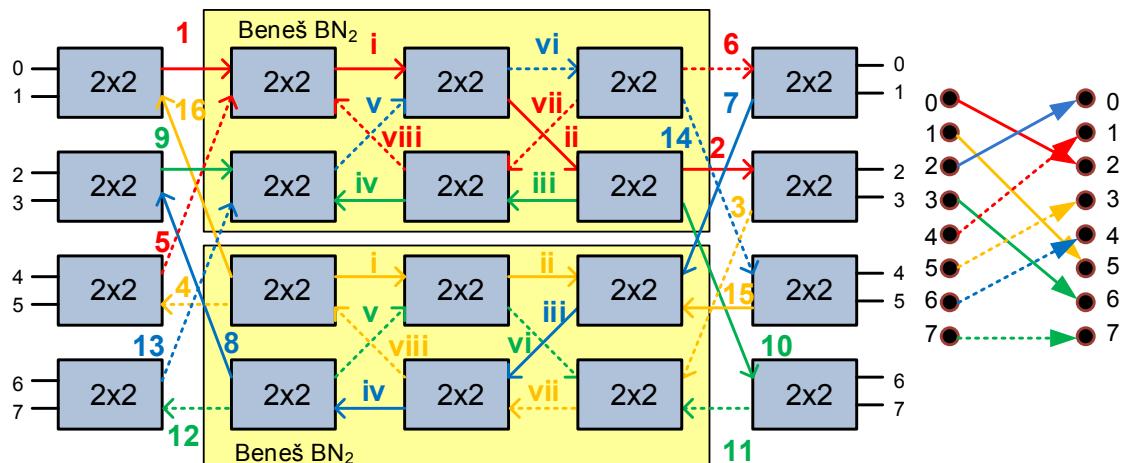
- N = 8 vstupů, 5 stupňů
- Použito např. u směrovače Cisco CSR-1.



Obrázek 18.20: Příklad Benešovy sítě.

### Algoritmus hledání propojení

- Horní podsíť BN<sub>i</sub> slouží pro dopředné směrování, spodní podsíť BN<sub>i</sub> pro zpětné.
- Postupuje se od vstupů k výstupům a zpět, vždy pro sousední porty.
- Nejprve se propojí vstupní a výstupní bloky, poté rekursivně síť BN<sub>i-1</sub>.



Obrázek 18.21: Příklad Benešovy sítě s přeskládáním.

# Kapitola 19

## PDS – Základní funkce směrovače, zpracování paketů ve směrovači, typy přepínání a architektur.

### 19.1 Zdroje

- 05-routing.pdf
- PDS\_2021-03-12.mp4

### 19.2 Úvod a kontext

**Fragmentace paketů** Uzel v síti (směrovač) dostane paket o velikosti  $n$ . Paket má přepo-  
slat přes výstupní rozhraní do sítě, ve které je  $MTU < n$ . Aby paket mohl být odeslán,  
musí být rozdělen (fragmentován) na více menších paketů (fragmenty) a odeslán po čás-  
tech. Na straně příjemce pak musí nastat opačný proces – defragmentace.

**MTU** MTU (*Maximum Transmission Unit*) je největší velikost paketu, kterou lze v síti  
odeslat (přes výstupní rozhraní sítového prvku). Závisí na sítové kartě příslušného roz-  
hraní.

### 19.3 Základní popis směrovače

Směrovač je sítové zařízení, které předává datové pakety mezi počítačovými sítěmi. Zá-  
kladní charakteristika:

- Pracuje s pakety (sítová vrstva, L3).
- Tvoří rozhraní mezi počítačovými sítěmi (provádí překlad NAT).
- Klasifikuje a filtruje pakety (firewall, ACL).
- Provádí fragmentaci a defragmentaci.

### Činnost

1. Vypouzdření paketu z L2 (odebrání L2 hlavičky) a kontrola jestli je v pořádku (po-  
mocí kontrolního součtu).

2. Vyhledání cesty kam se má paket směrovat a překlad adresy NAT (pomocí směrovací tabulky).
3. Určení cílové MAC adresy na základě cílové IP adresy (pošle ARP dotaz).
4. Určení výstupního rozhraní.
5. Sestavení výsledného paketu podle výstupního rozhraní (zapouzdření do příslušné L2 technologie – přidání L2 hlavičky).

### Co ovlivňuje propustnost

- Rozbalení paketu.
- Vyhledání směrovací cesty.
- Překlad NAT.
- Vyhledání cílové MAC adresy.
- Zabalení paketu do správné technologie.

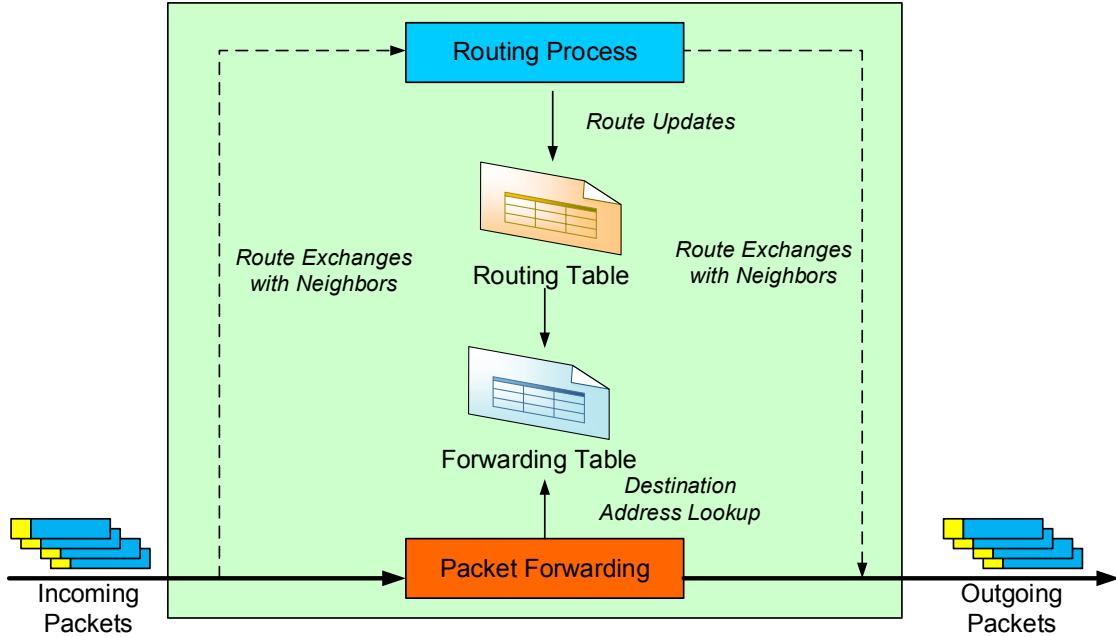
### Typy

- Páteřní směrovače – součástí pátečních sítí mezi ISP.
- Hraniční směrovače – mezi zákaznickými sítěmi a ISP.
- Podnikové směrovače – propojení koncových systému v podnikových (*enterprise*) sítích.
- Domácí směrovače.

## 19.4 Operace co směrovač vykonává

Pakety mohou být určeny buď přímo pro směrovač (aktualizace směrovací cesty přes směrovací protokoly – aktualizace směrovací tabulky) a nebo jinému uzlu v síti. V takovém případě směrovač paket přeposílá dalším uzlům v síti.

- **Směrování (routing)** – Určování cest paketů v prostředí počítačových sítí. Cílem je doručit paket adresátovi, pokud možno co nejefektivnější cestou. Směrování zajišťují nejen směrovače, ale i koncové stanice (při vysílání).
- **Přepínání (forwarding, switching)** – Přepnutí paketu ze vstupního rozhraní na výstupní (viz architektura přepínačů).



Obrázek 19.1: Základní činnost směrovače.

**Směrovací tabulka** Směrovací tabulka (*routing table*) obsahuje informace nutné a dostávající pro směrování (ipprefix – IP prefix cílové sítě, nexthop – IP adresa dalšího uzlu, metrika – cena cesty, počet hopů). Informace do ní se získavají pomocí směrovacích protokolů a nebo staticky (administrátor provede manuálně). Směrovací tabulku mají i koncové stanice (počítače), aby mohly posílat pakety do internetu.

| IP prefix    | Next hop      |
|--------------|---------------|
| 10.5.0.0/16  | 192.168.2.254 |
| 10.15.0.0/16 | 104.17.2.1    |
| 88.0.0.0/8   | 129.1.1.1     |

Obrázek 19.2: Příklad směrovací tabulky (bez metriky).

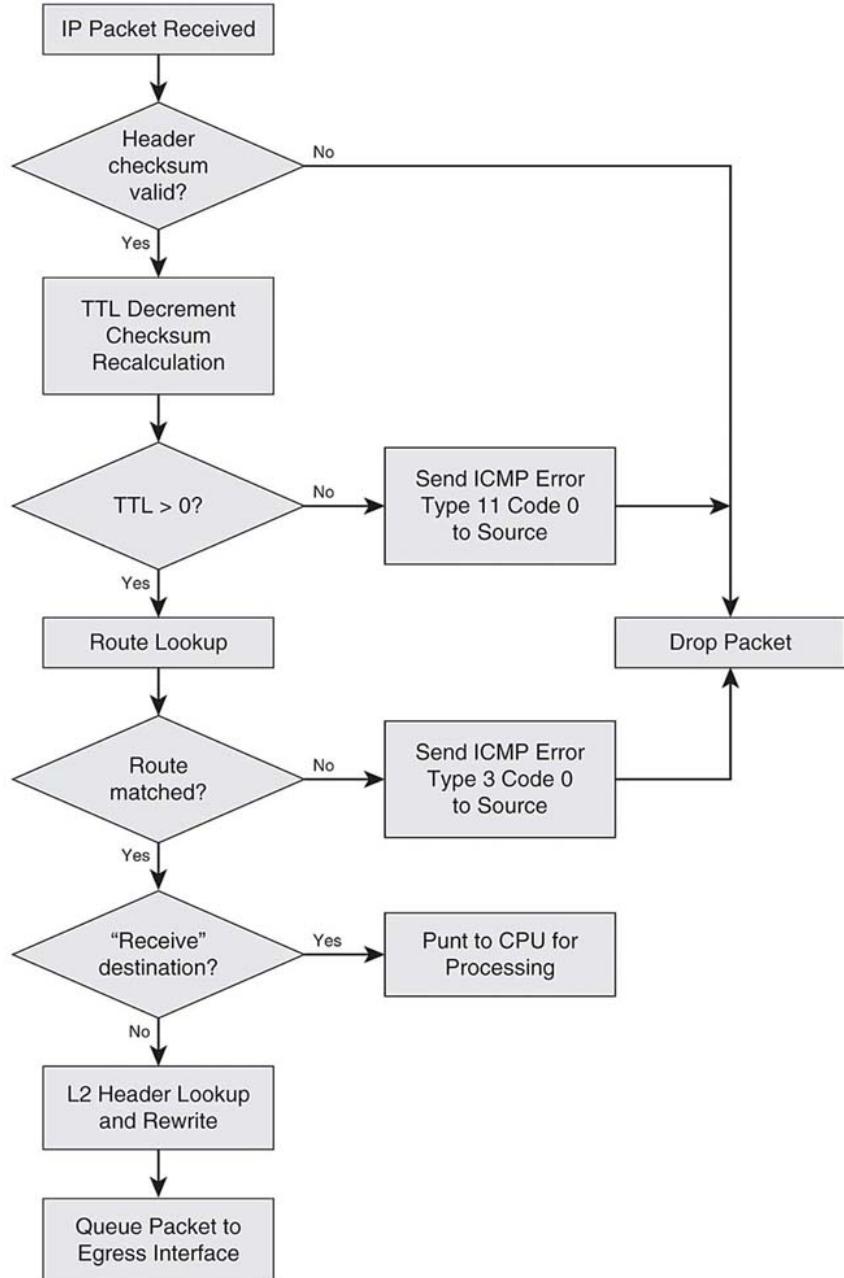
**Přepínací tabulka** Přepínací tabulka (FIB, *forwarding information base, forwarding table*) doplňuje směrovací tabulku o další informace, které jsou nutné pro sestavení cílového paketu (výstupní rozhraní, zdrojová MAC adresa, cílová MAC adresa). Informace do ní se získavají pomocí ARP protokolu a doplnění vlastních informací (*dst MAC address*).

| IP prefix    | Interface | dst MAC address   | src MAC address   |
|--------------|-----------|-------------------|-------------------|
| 10.5.0.0/16  | eth0      | 00:0F:1F:CC:F3:06 | B8:6B:23:EA:FC:20 |
| 10.15.0.0/16 | eth1      | 01:12:11:A0:17:A0 | 00:16:17:E1:28:5F |
| 88.0.0.0/8   | eth2      | 01:3F:04:10:03:15 | 00:1B:63:8A:DB:1A |

Obrázek 19.3: Příklad přepínací tabulky.

## Zpracování IP paketu (již vypouzdřeno z L2)

1. Validace hlavičky L3 (formát, verze, délka).
2. Kontrola hodnoty TTL a její dekrementace. Pokud je TLL 0, tak se paket zahodí a pošle ICMP zpráva.
3. Přepočítání kontrolního součtu.
4. Zpracování rozšířených voleb IP protokolu (timestamp, record route, strict source route).
5. Vyhledání cesty (next-hop) na základě cílové adresy (lokální doručení, unicast, multicast). Pokud se nepodaří vyhledat cestu, tak se paket zahodí a pošle ICMP zpráva.
6. Fragmentace a defragmentace IP paketů (pokud  $MTU_{in} < MTU_{out}$ ).
7. Zpracování zpráv ICMP a IGMP.



Obrázek 19.4: Diagram zpracování paketu ve směrovači.

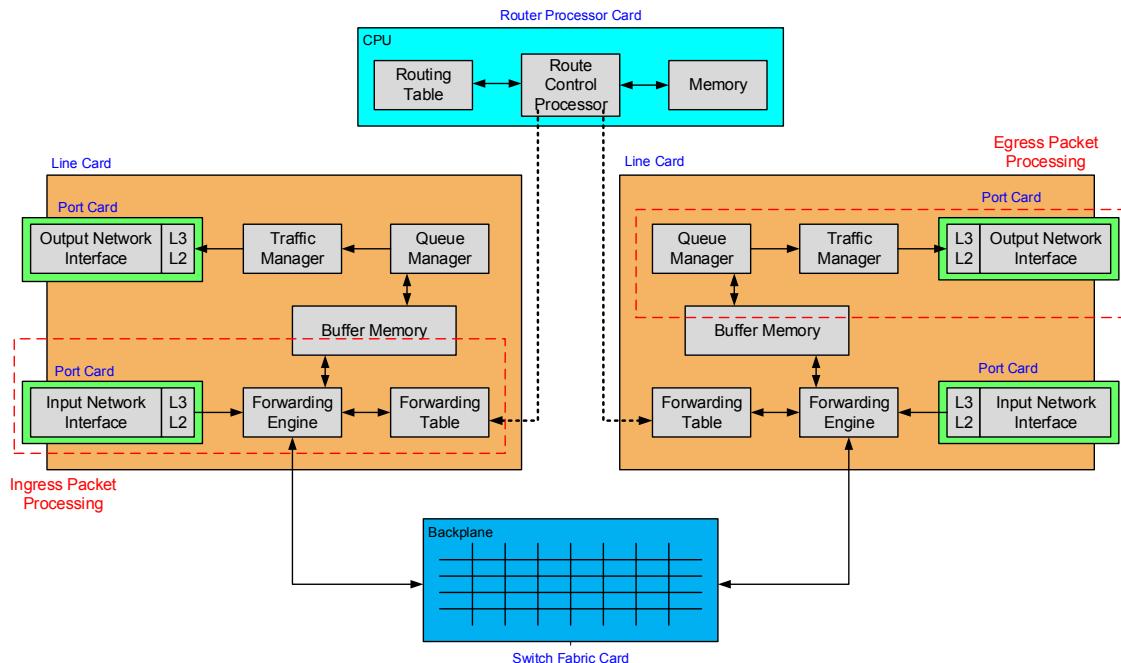
## 19.5 Architektura směrovače

Rozdělíme na fyzické části (z hlediska hardware) a funkční části (z hlediska komponent co něco vykonávají).

### 19.5.1 Fyzické části

- Procesorová část (*Router Processor Card*)
- Přepínací logika (*Switch Fabric Card*)

- Sít'ové karty (*Line Card*) – každá má vstupní a výstupní rozhraní (*Port Card*)



Obrázek 19.5: Fyzické části (procesory, paměti) směrovače.

### 19.5.2 Funkční části

**Procesorová část** procesor, paměť, směrovací tabulka

- Implementuje směrování na obecném CPU
- Zpracovává směrovací informace: aktualizace, udržuje sousedství
- Obsluhuje směrovací tabulkou
- Přenáší data do přepínací tabulky
- Zpracovává pakety, které nelze směrovat pomocí FIB
- Vytváří chybové zprávy ICMP

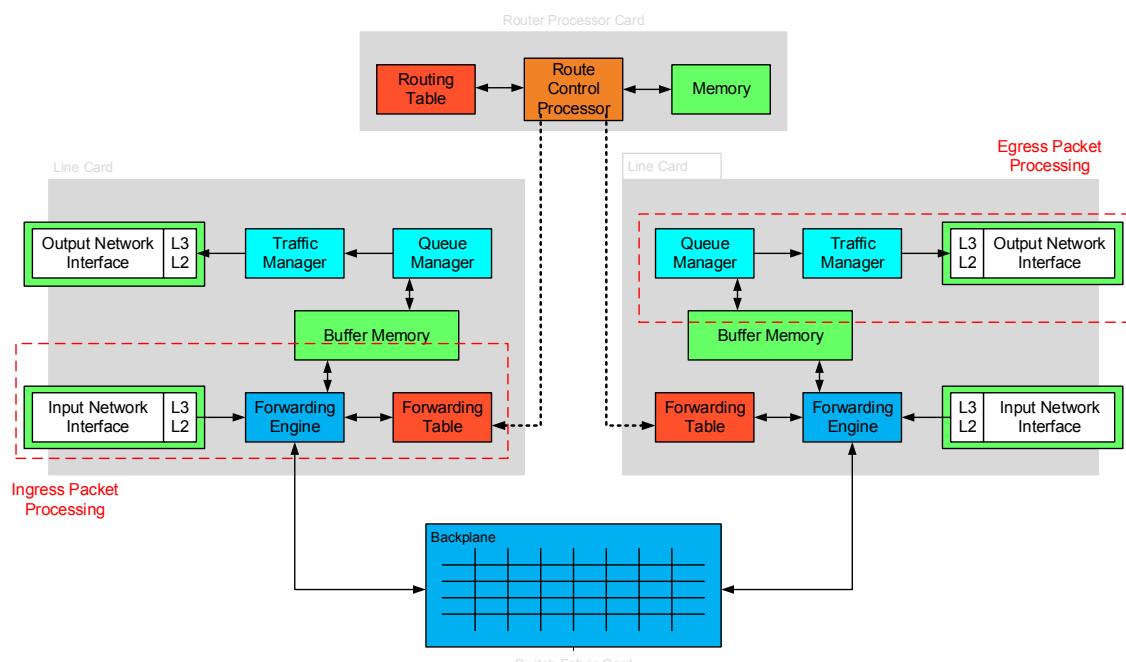
**Přepínací logika** propojovací deska (*Backplane*)

- Propojuje sít'ové rozhraní
- Vytváří sdílené (shared) či přepínané (switched) propojení
- Rychlosť přepínání odpovídá přenosovému pásmu všech rozhraní

**Sít'ové karty**

- Vstupní/Výstupní sít'ové rozhraní (*Input/Output Network Interface*)
  - Obsahuje více vstupních/výstupních portů
  - Odstraní zapouzdření L2
  - Předá hlavičku L3 přepínacímu modulu

- Uloží paket do paměti
- Zapouzdří odchozí pakety
- Řízení přepínání (*Forwarding Engine*)
  - Dostane hlavičku L3 ze síťového rozhraní
  - Určí výstupní rozhraní podle informace ve FIB
  - Provádí klasifikaci paketů pro podporu QoS na výstupu
- Správce front (*Queue Manager*)
  - Ukládá pakety do vyrovnávací paměti, pokud je výstupní port obsazen
  - Spravuje výstupní frontu → různé typy výstupních front
  - Při zaplnění fronty vybírá a zahazuje pakety podle definované politiky
- Správce provozu (*Traffic Manager*)
  - Prioritizuje a reguluje výstupní provoz podle požadavků QoS
  - Omezuje či ořezává výstupní provoz (shaping, policing)

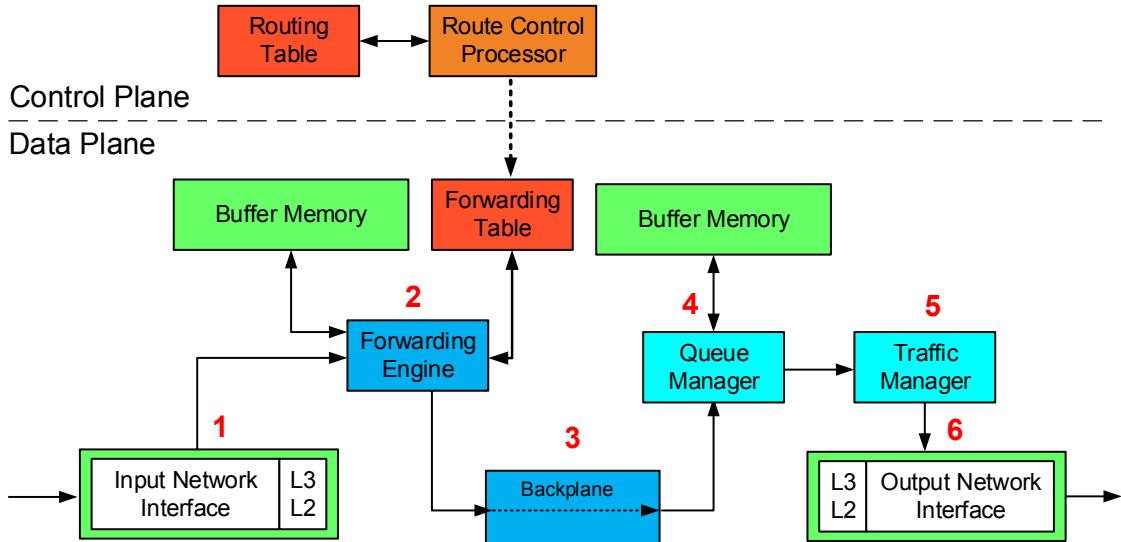


Obrázek 19.6: Funkční částí (procesy).

## 19.6 Zpracování paketu ve směrovači

Směrovač můžeme rozdělit na dvě části:

- *Data Plane* – Specifický hardware pro zpracování paketu (rychlé zpracování) – ASIC.
- *Control Plane* – Obecný procesor pro realizaci složitějších operací (pomalé zpracování).



Obrázek 19.7: Rozdělení přepínače na části *Data Plane* a *Control Plane* a zpracování rozděleno do 6. fází.

**Kontext paketu** Pomocná datová struktura, která reprezentuje paket uvnitř směrovače. Obsahuje vybrané informace z hlaviček. V rámci jednotlivých komponent směrovače se předává (pokud možno) tato struktura, pro větší efektivitu. Na počátku máme pouze částečný kontext, který se průběžně doplňuje. Na základě něho je pak sestaven výstupní paket.

**Fáze zpracování** Pokud některá z fází skončí neúspěšně, paket je zahozen.

1. Paket přijde na síťové rozhraní
  - Sítová karta zpracuje L2 rámcem, zkонтroluje FCS
  - Vytvoří kontext paketu: vloží L2 zdrojovou a cílovou adresu
  - Zpracování hlavičky L3: typ protokolu, kontrolní součet, TTL
  - Doplnění kontextu o informace L3: IP adresy, typ protokolu, DSCP + porty
2. Zpracování v přepínacím modulu
  - Vyhledání cesty v přepínací tabulce: next hop + výstupní rozhraní
  - Doplnění dalších informací do kontextu paketu
  - Paket uložen do vyrovnávací paměti → adresa vložena do kontextu
3. Přeposlání kontextu propojovací deskou
  - Paket i kontext jsou přeneseny na výstupní rozhraní
4. Zpracování správcem front
  - Podle priority v kontextu paketu je paket vložen do příslušné výstupní fronty
  - Obsluha fronty probíhá podle daného plánovacího algoritmu
5. Předání kontextu správci provozu
  - Kontrola omezení rychlosti (shaping) dle kontextu

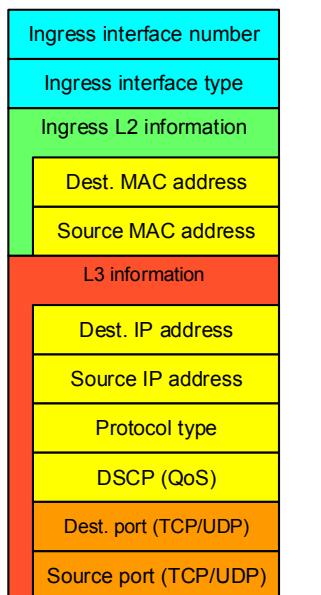
- Překročení rychlost: zahození či zpomalení

## 6. Výstupní síťové rozhraní

- L3: aktualizace TTL, přepočítání kontrolního součtu
- L2: přidání hlavičky, výpočet CRC
- Odeslání paketu (zapsání na výstupní médium)

### Plný kontext paketu

- Kontext doplněn o výstupní informace
- Základní deska přenese paket i kontext na výstupní rozhraní
- Kontext obsahuje adresu uložení paketu v paměti
- Zpracování paketu předáno správci front



(a) Kontext paketu – částečný.



(b) Kontext paketu – úplný.

Obrázek 19.8: Kontext paketu (*ingress* – vstup, *egress* – výstup.)

Paket může být zahozen i na výstupní kartě. Proč?

- Vypršelo TTL
- Jsou plné fronty (politika RED, WRED)
- Filtrování na výstupu

Paket může být zpracován dvěma způsoby: rychlou cestou (přes *data plane*) a pomalou cestou (přes *control plane*).

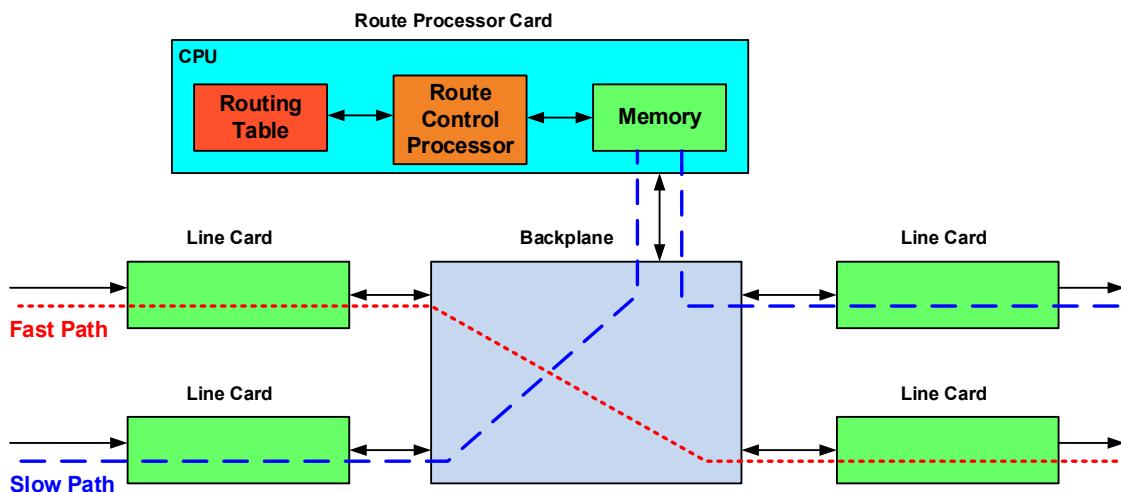
**Rychlá cesta** Paket je možné zpracovat pouze v hardwaru (*data plane*).

- Zpracování hlavičky IP (kontrola verze, délky paketu, snížení TTL, přepočítání kontrolního součtu)

- Přeposlání paketu ze vstupní na výstupní (lokální uložení, přeposlání na jeden port (unicast) či více portů (multicast))
- Klasifikace paketu na základě informací z hlavičky (optimalizované datové struktury pro rychlé uložení a vyhledání)
- Uložení do front, plánování (různé typy front, různé typy obsluhy)

**Pomalá cesta** Paket je částečně zpracován v hardwaru, ale většina zpracování musí proběhnout softwarově (*control plane*). Síťová karta nedokáže zpracovat paket sama a musí se využít centrální procesor.

- Zpracování ARP (zjištění výstupní adresy L2: první paket, ostatní pakety)
- Fragmentace a defragmentace
- Generování zpráv ICMP
- Zpracování směrovacích informací



Obrázek 19.9: Dva způsoby zpracování paketu ve směrovači – pomalou cestou a rychlou cestou.

## 19.7 Typy přepínání paketů

Přepínání paketů z jednoho rozhraní na druhé na základě směrovacích informací je jedna z nejdůležitějších funkcí směrovače. Proces přepínání paketů zahrnuje:

- Zjištění, zda cíl cesty paketu je dosažitelný.
- Vyhledání nejbližšího uzlu na cestě (next-hop) a určení výstupního rozhraní.
- Vyhledání informací pro vytvoření L2 hlavičky paketu na výstupu.

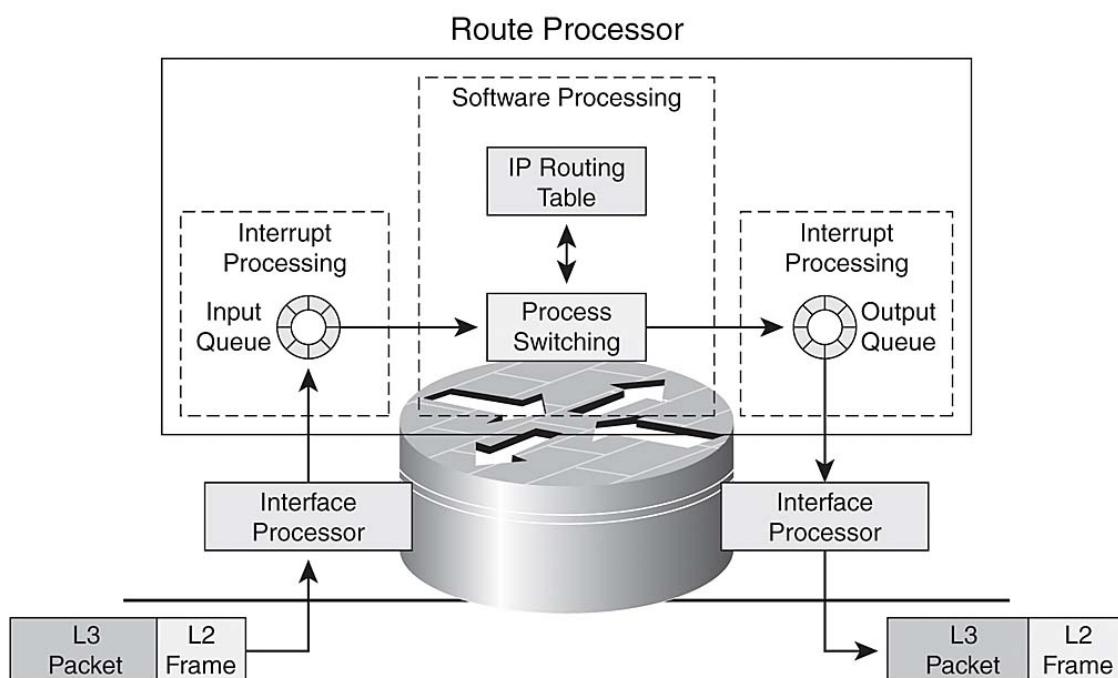
### 19.7.1 Softwarové přepínání (*Process Switching*)

- Přepínání pomocí centrálního procesoru s centrální pamětí.
  - Pomalé – přístup do paměti, obecný procesor (softwarové zpracování).
- Pro každý paket se hledá cesta ve směrovací tabulce a určuje se MAC adresa cíle.

- Směrovač využívá standardní mechanismus přepíná procesů v OS (přeřušení).

### Fáze softwarového přepínání

1. I/O procesor detekuje paket na vstupním médiu. Přenese ho do vstupního bufferu.
2. I/O procesor vygeneruje přerušení. Během přerušení určí centrální procesor typ paketu a zkopíruje ho do centrální paměti.
3. Centrální plánovač zjistí, že ve vstupní frontě je paket. Naplánuje jeho další zpracování procesem.
4. Proces pro zpracování vyhledá ve směrovací tabulce další uzel (next hop) a výstupní rozhraní. V paměti ARP cache vyhledá MAC adresu dalšího uzlu.
5. Přepíše L2 hlavičku paketu a umístí paket do výstupní fronty na výstupním portu.
6. Paket vložen do fronty na výstupním portu.
7. I/O procesor detekuje paket ve vysílací frontě. Zapíše ho na síťové médium.



Obrázek 19.10: Softwarové přepínání.

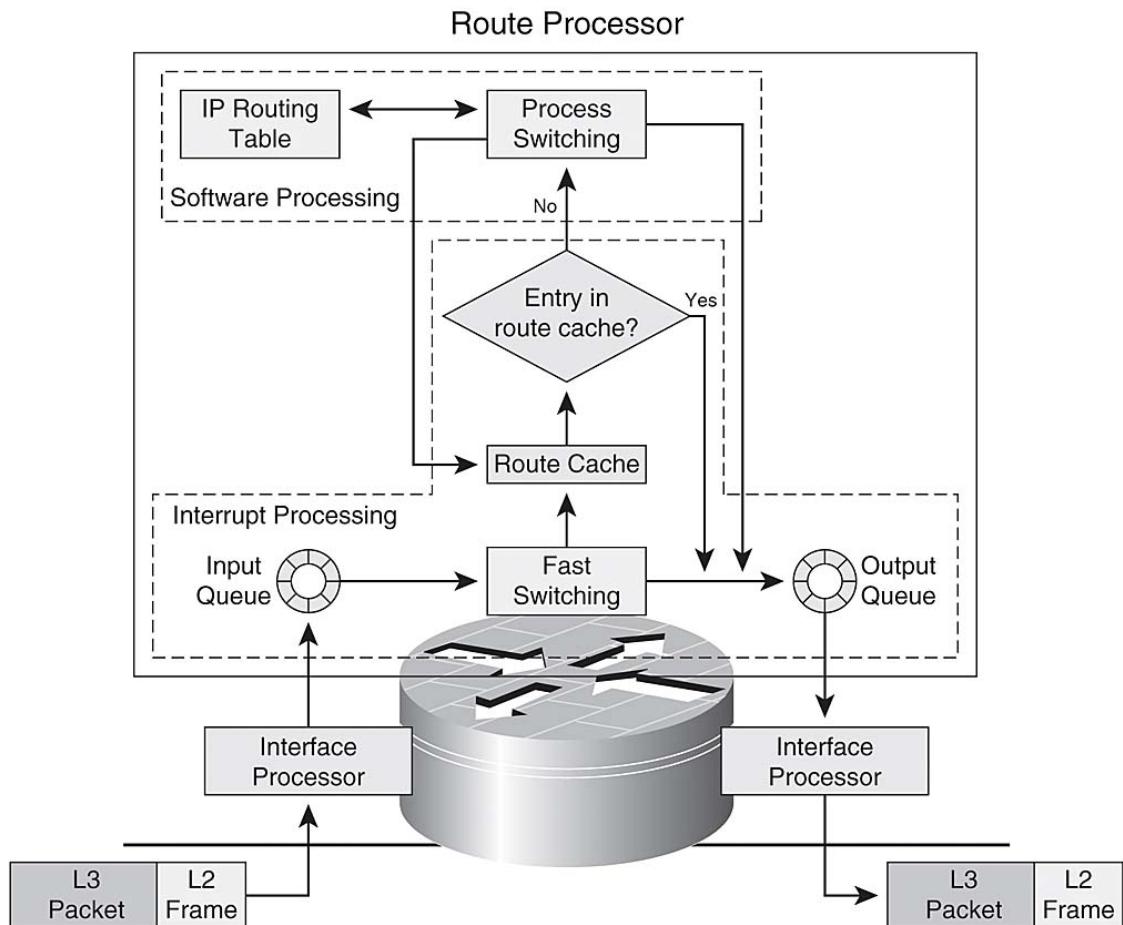
### Optimalizace

- Nejdéle trvá: na základě cíle vyhledávání informací ve směrovací tabulce, najdu next hop, najdu příslušný záznam v ARP tabulce
- Můžeme využít cache (která si bude pamatovat poslední záznamy).

#### 19.7.2 Rychlé přepínání (*Fast Switching*)

- Využívá vyrovnávací paměť *route cache* s předpočítanou L2 hlavičkou.
- První paket toku se přepíná softwarově, další pakety toku jdou rychlou cestou.

- Při přepínání paketu je vložen do paměti *route cache* nový záznam.
- Neexistuje synchronizace mezi směrovací tabulkou, ARP cache a *route cache*.
- Záznam v *route cache* se zneplatní při změně ARP cache či směrovací tabulky.
- Při zaplnění paměti nad určitou mez se začnou záznamy náhodně zahazovat.

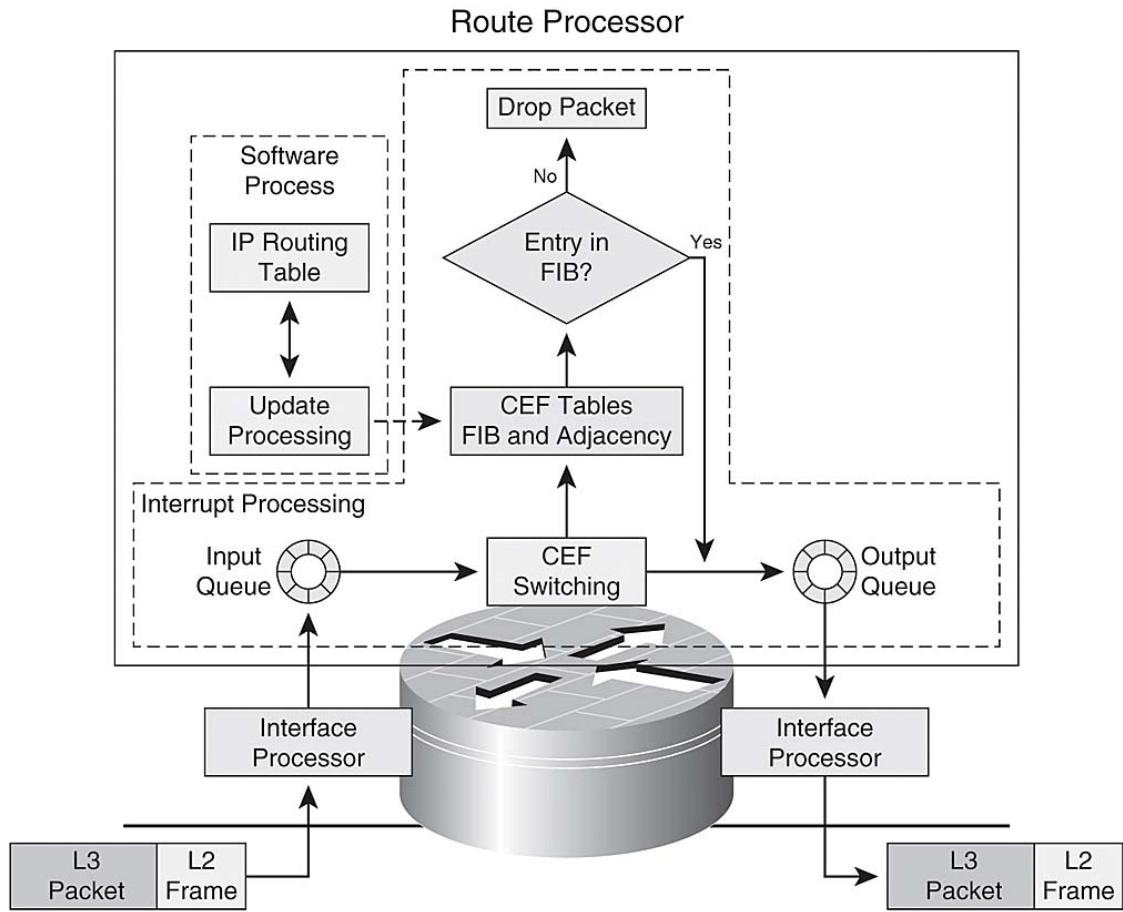


Obrázek 19.11: Rychlé přepínání.

### 19.7.3 Expresní přepínání CEF (Cisco Express Forwarding)

[[todo]]

- Tabulka CEF se předpočítá na základě směrovací tabulky a tabulky sousedů ještě před příchodem paketu → nedochází k softwarovému přepínání.
- Oddělení směrovacích informací od L2 dat → nedochází ke stárnutí záznamů při expiraci záznamu v tabulce ARP.
- Změny ve směrovací tabulce či tabulce ARP se okamžitě propagují do tabulky CEF.
- ARP tabulka se synchronizuje se záznamy v tabulce sousedů.



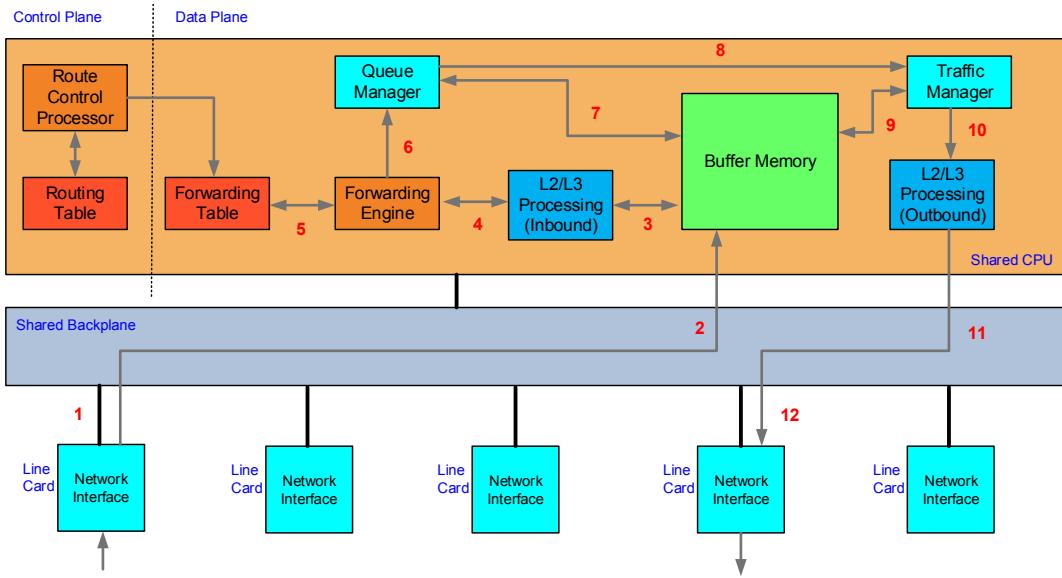
Obrázek 19.12: Expresní přepínání CEF.

## 19.8 Přehled architektur

Architektury směrovačů rozdělmě podle způsobu přepínání paketů.

### 19.8.1 Architektura se sdíleným procesorem

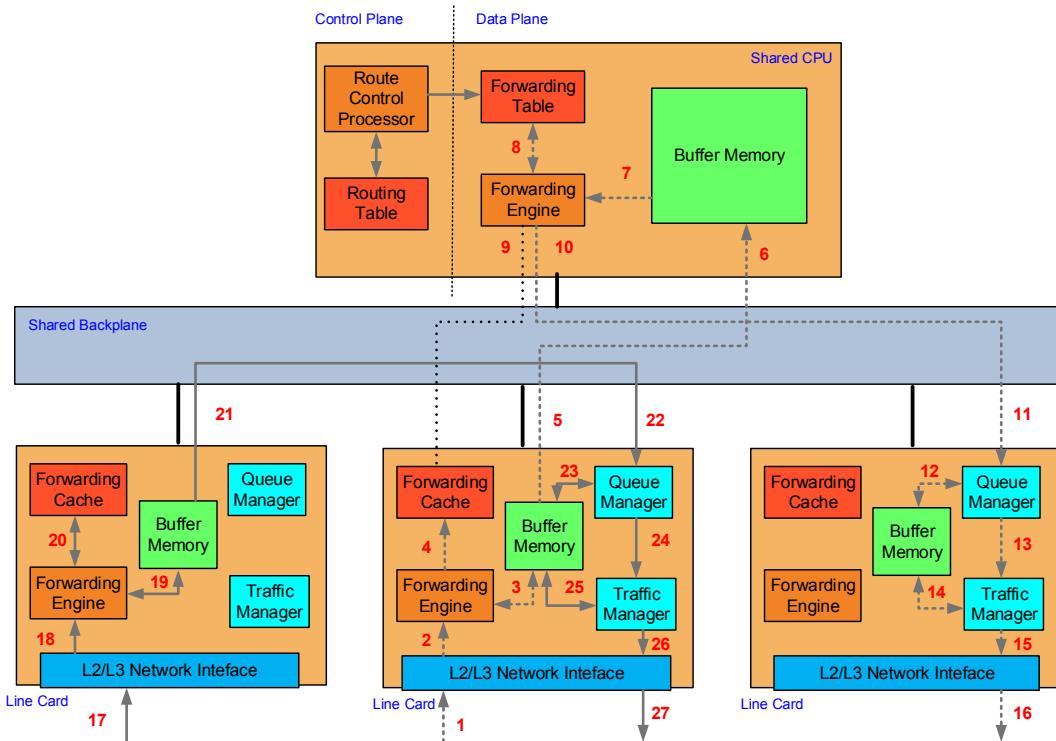
- Využívá softwarové přepínání → každý paket zpracován na CPU.
- Cykly CPU rozděleny mezi přepracování, směrování a další operace.
- Sdílená sběrnice i procesor ⇒ levné, ale pomalé.



Obrázek 19.13: Architektura se sdíleným procesorem.

### 19.8.2 Architektura se sdíleným procesorem a pamětí cache na kartě

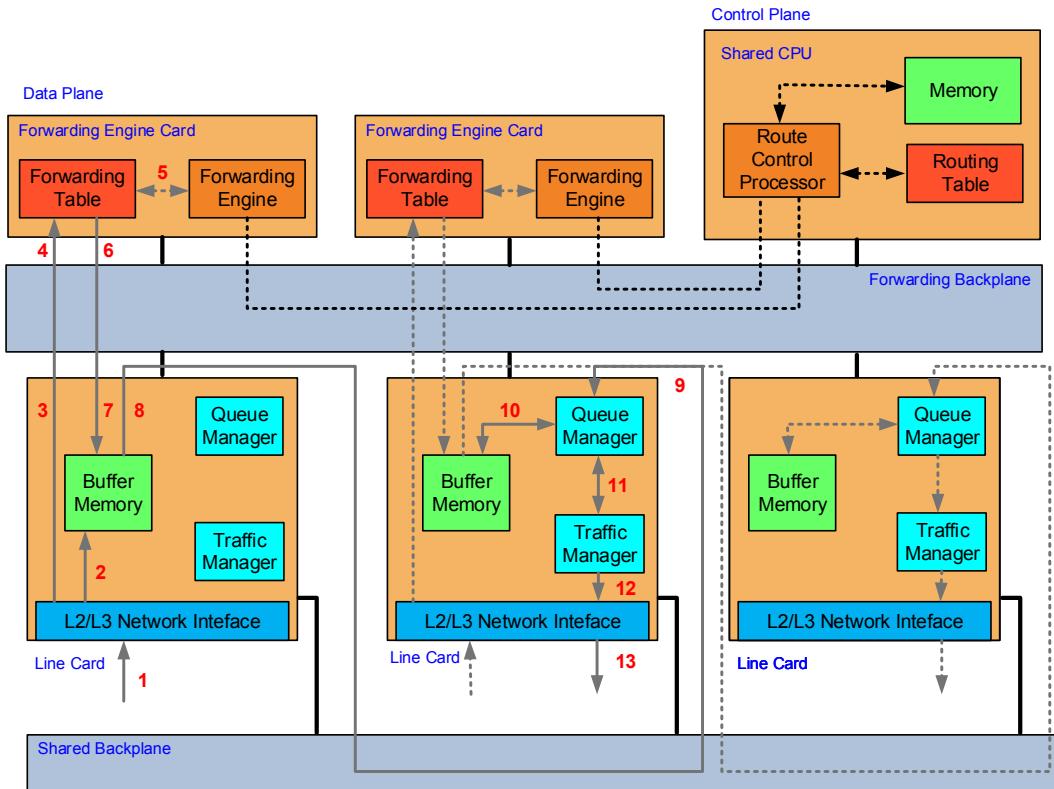
- Varianta s pamětí cache na kartě → synchronizace přepínacích tabulek.
- Sítová karta obsahuje FE pro zpracování hlaviček, paměť a přepínací tabulkou.
- Rychlé přepínání (Fast Switching): první paket vs. další pakety.



Obrázek 19.14: Architektura se sdíleným procesorem a pamětí cache na kartě.

### 19.8.3 Architektura s nezávislými moduly FE

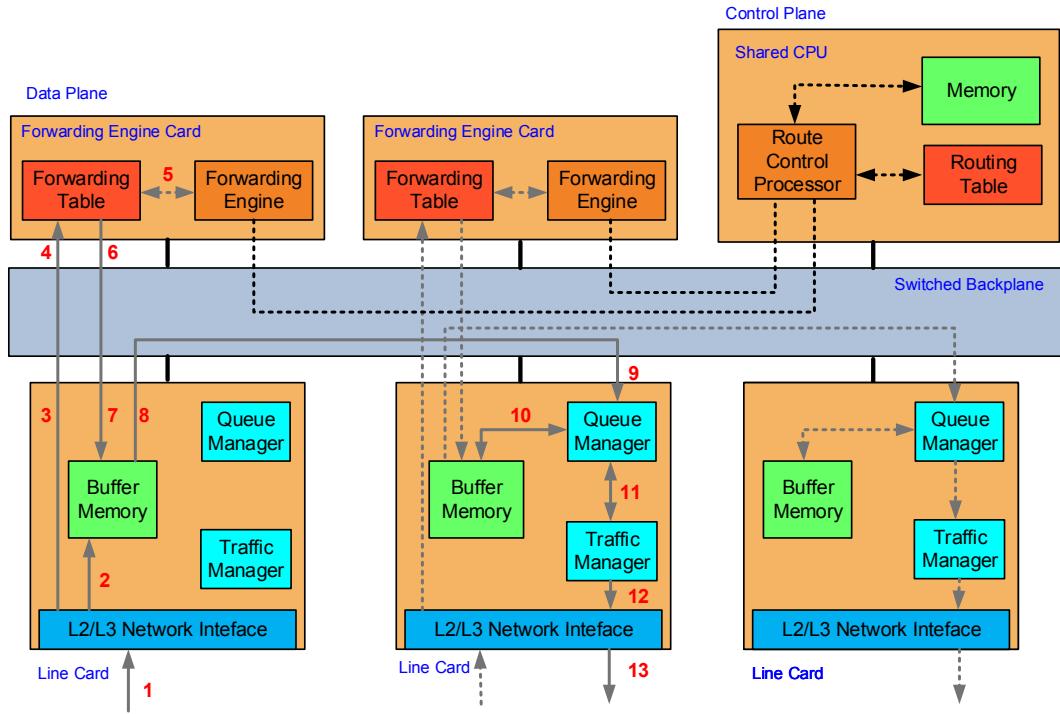
- Přepínací moduly FE implementovány na speciálních kartách.
- Paralelní zpracování paketů, dvě sběrnice (sdílená a přepínaná).



Obrázek 19.15: Architektura s nezávislými moduly FE.

### 19.8.4 Architektura s nezávislými moduly FE a přepínanou sběrnicí

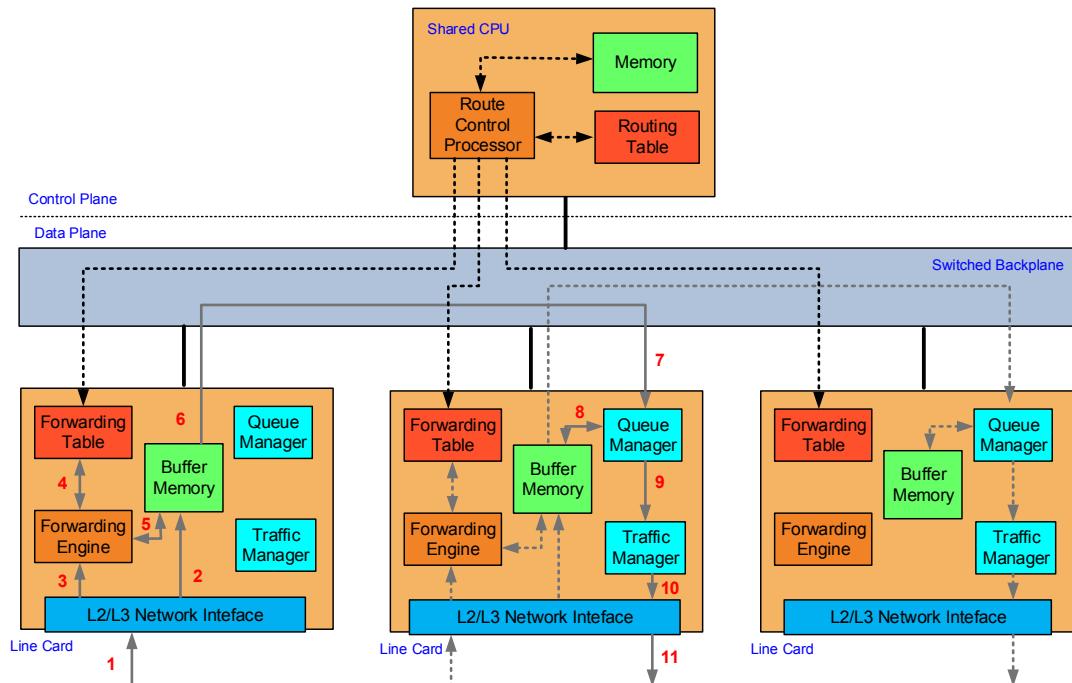
- Varianta s jednou přepínanou sběrnicí ⇒ vyšší propustnost.



Obrázek 19.16: Architektura s nezávislými moduly FE a přepínanou sběrnicí.

### 19.8.5 Distribuovaná architektura (Shared Nothing)

- Veškeré zpracování paketu přeneseno do síťového modulu.
- Oddělení procesu směrování a přepínání → využití technologie CEF.



Obrázek 19.17: Distribuovaná architektura (Shared Nothing).

# Kapitola 20

## PDS – Metody pro výpočet směrování v sítích (Bellman-Ford, Dijkstra, Path vector, DUAL).

### 20.1 Zdroje

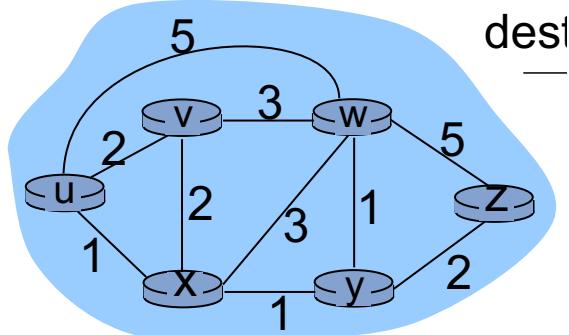
- 09-teorie-smerovani.pdf
- PDS\_2021-04-16.mp4

### 20.2 Směrování

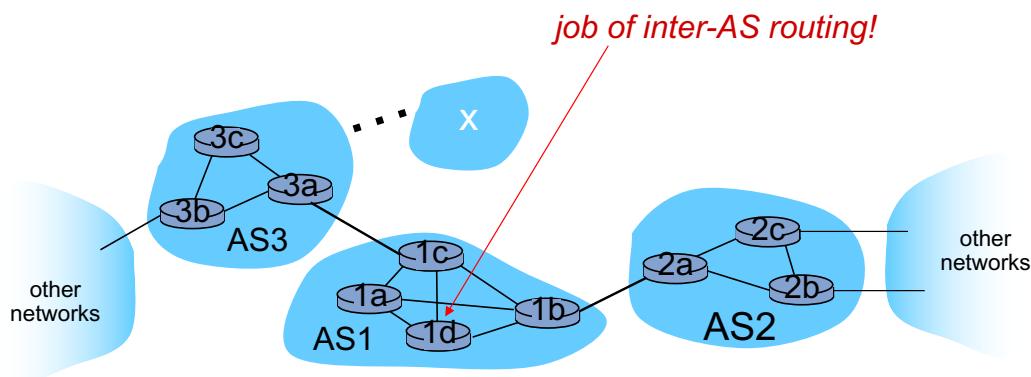
Z pohledu směrování můžeme počítačovou síť zjednodušit na neorientovaný, ohodnocený graf, kde uzly jsou směrovače a hrany jsou propojení mezi něma. Směrování je pak hledání nejkratší cesty mezi dvěma uzly, resp. ze zdrojového do všech.

**Autonomní systém** Není možné směrovat přes celý internet (příliš velký, příliš mnoho uzlů). Využíváme abstrakce, kdy se na podsíť díváme jako na uzel – tzv. autonomní systém. Tímto agregujeme více informací do jedné (přesné adresy uzlů, na vyšší úrovni reprezentujeme pouze adresou sítě). Internet je pak sítí sítí. Může být více úrovní (autonomní systém autonomních systémů).

**Hierarchické směrování** Hierarchické směrování znamená, že nesměrujeme mezi každým směrovačem na světě, ale využíváme autonomní systémy. Směrujeme tedy globálně mezi autonomníma systémama a pak uvnitř mezi směrovačema.



Obrázek 20.1: Příklad počítačové sítě zobrazené jako neorientovaný ohodnocený graf  
 $G = (V, E)$ ,  $V = \{u, v, w, x, y, z\}$ ,  $E = \{(u, v), \dots\}$ ,  $w = \{((u, v), 2), \dots\}$ .



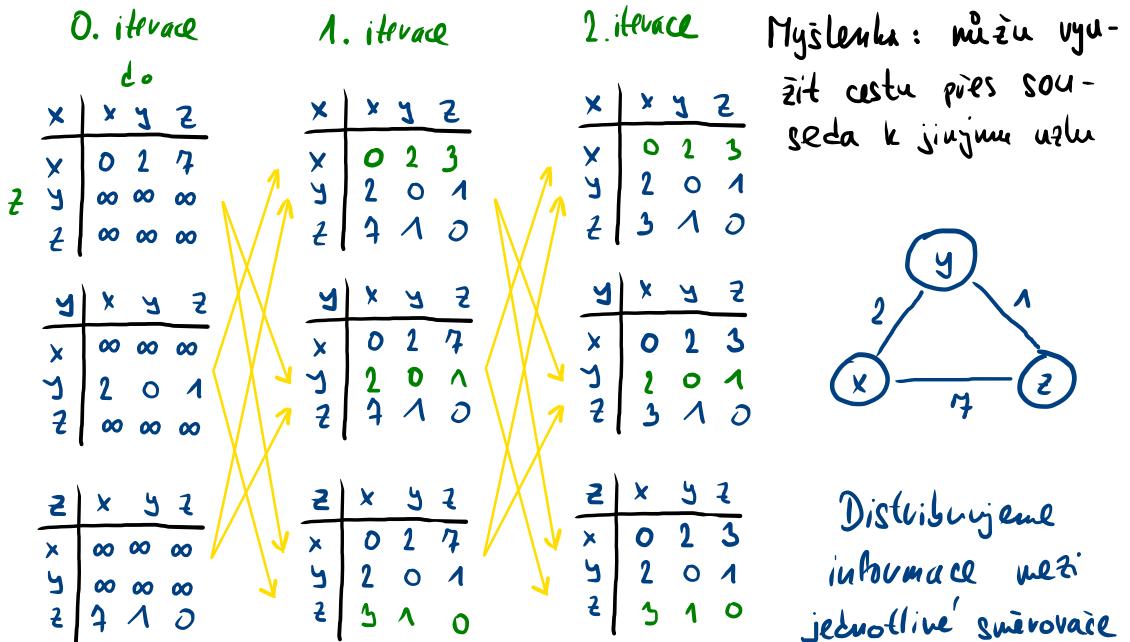
Obrázek 20.2: Příklad autonomních systémů (např. VUT, CESNET, UPC, ...).

## 20.3 Distance Vector přístup

- Distribuované směrovací informace – Každý uzel, zná pouze omezenou část topologie sítě. Konkrétně má informace pouze co sám zná a informace od svých sousedů.
- Používá se pro směrování uvnitř autonomních systémů.
- *Single-metric*.

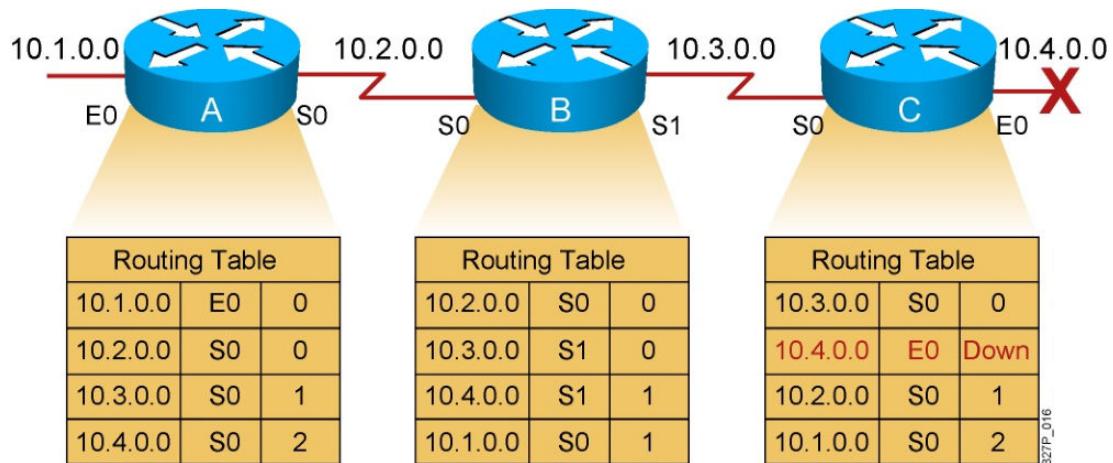
### 20.3.1 Algoritmus Bellman-Ford

- Pro formální vysvětlení a pseudokód viz otázku „Hledání nejkratších cest ze zdrojového uzlu do všech ostatních uzlů grafu“.
- Jde o algoritmus hledání nejkratších cest ze zdrojového uzlu do všech ostatních uzlů. To znamená, že ho provádí každý uzel (směrovač).
- Využívá jej např. směrovací protokol RIP (*Routing Protocol*).
  - Komunikuje přes UDP na portu 520.
  - Metrika: *hop count*.
  - Proti *Infinity Count* (viz dále) se brání vylepšením *Split Horizon*: Nikdy se nepošle informace o cestě zpátky rozhraním, ze kterého přišla.



Obrázek 20.3: Příklad výpočtu nejkratší cesty pomocí algoritmu Bellman-Ford.

**Infinity Count** V případě, že vypadne některá ze sítí, může nastat problém počítání ceny cesty do nekonečna. Jelikož si uzly vyměňují pravidelně směrovací informace. Viz následující obrázek.

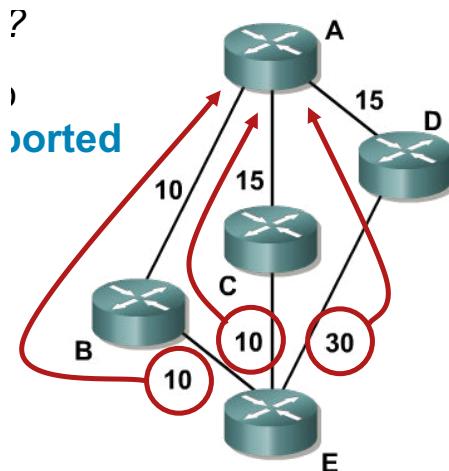


Obrázek 20.4: Příklad *Infinity Count*. Síť 10.4.0.0 vypadne a pro uzel C je najednou nedosažitelná. Uzel B se cestu do sítě 10.4.0.0 naučil dříve od C. Uzel C čerpá informaci o cestě do sítě z uzel B. Tímto způsobem se cena do sítě zacyklí a jde k  $\infty$ .

### 20.3.2 Diffusing update algorithm (DUAL)

- Algoritmus DUAL (diffusing update algorithm) je algoritmus společnosti Cisco, který zajišťuje, že daná trasa je přeypočítána globálně, kdykoli by mohla způsobit směrovací smyčku.

- Klíčová je tzv. **Feasibility Condition**, která zajišťuje, že jsou vždy vybírány pouze trasy bez smyček. Pokud podmínka platí, nemohou vzniknout žádné smyčky, ale podmínka může za určitých okolností odmítnout všechny trasy k cíli, přestože některé jsou bez smyček.
  - Vzdálenost sousedů k cílové síti nazveme *RC* (*reported distance*).
  - Nejlepší vzdálenost k danému uzlu nazveme *FD* (*feasible distance*).
  - Pokud platí  $RD < FD$ , tak cesta neobsahuje smyčku.
- Využívá jej směrovací protokol EIGRP (*Enhanced Interior Gateway Routing Protocol*) od společnosti Cisco.
  - Pro hledání nejkratší cesty používá algoritmus Bellman-Ford + *Feasibility Condition*.
  - Komunikace: vlastní protokol zabalený do IP paketu (*Reliable Transport Protocol*).
  - Kompozitní metrika – jedno číslo spočítaný na základě několika parametrů (šířka pásma, zpoždění, rychlosť, ...).



Obrázek 20.5: Příklad výpočtu nejlepší cesty a uplatnění feasibility condition.

Cesta z A do E:

$$\begin{aligned} RD(B) &= 10 & RD(C) &= 10 & RD(D) &= 30 \\ via(B) &= 20 & via(C) &= 25 & via(D) &= 45 \end{aligned}$$

Potom z pohledu A:

- Cesta  $via(B)$  je *FD* a platí podmínka  $FD > RD(B)$ , tudíž cesta přes B neobsahuje smyčku.
- Pro cestu  $via(C)$  platí podmínka  $FD > RD(C)$ , tudíž cesta přes C neobsahuje smyčku.
- Pro cestu  $via(D)$  neplatí podmínka  $FD > RD(D)$ , tudíž cesta přes D může obsahovat smyčku a není brána v potaz.

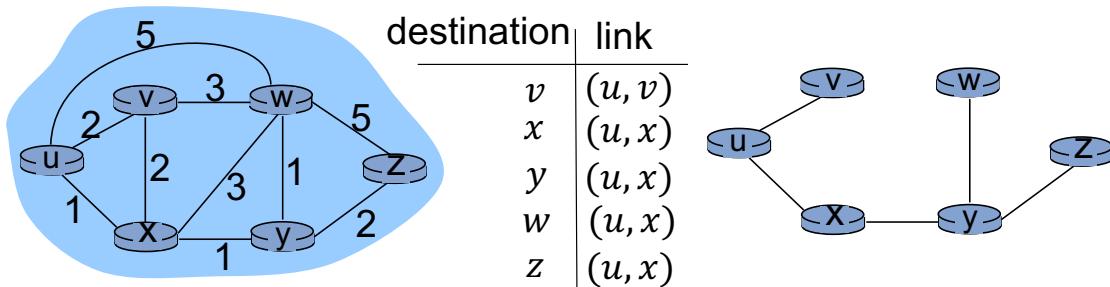
## 20.4 Link State přístup

- Globální směrovací informace – Každý uzel zná celou topologii sítě. Na začátku si uzly vymění informace o topologii sítě.
- Používá se pro směrování uvnitř autonomních systémů.
- *Single-metric.*

### Dijkstrův algoritmus

- Pro formální vysvětlení a pseudokód viz otázku „Hledání nejkratších cest ze zdrojového uzlu do všech ostatních uzlů grafu“.
- Jde o algoritmus hledání nejkratších cest ze zdrojového uzlu do všech ostatních uzlů. To znamená, že ho provádí každý uzel (směrovač).
- Používá ho např. směrovací protokol OSPF (*Open Shortest Path First*).
  - Komunikace: vlastní protokol nad IP.
  - Metrika: odvozena od rychlosti linky.

| Step | $N'$     | $D(v), p(v)$ | $D(w), p(w)$ | $D(x), p(x)$ | $D(y), p(y)$ | $D(z), p(z)$ |
|------|----------|--------------|--------------|--------------|--------------|--------------|
| 0    | $u$      | $2, u$       | $5, u$       | $1, u$       | $\infty$     | $\infty$     |
| 1    | $ux$     | $2, u$       | $4, x$       |              | $2, x$       | $\infty$     |
| 2    | $uxy$    | $2, u$       | $3, y$       |              |              | $4, y$       |
| 3    | $uxyv$   |              | $3, y$       |              |              | $4, y$       |
| 4    | $uxyw$   |              |              |              |              | $4, y$       |
| 5    | $uxyvwz$ |              |              |              |              |              |



Obrázek 20.6: Příklad výpočtu nejkratší cesty pomocí algoritmu Dijkstra. Step značí iteraci,  $N'$  je množina již prozkoumaných uzlů,  $D$  je pole vzdáleností do uzlu,  $p$  je pole předchůdzů uzlu.

## 20.5 Path Vector přístup

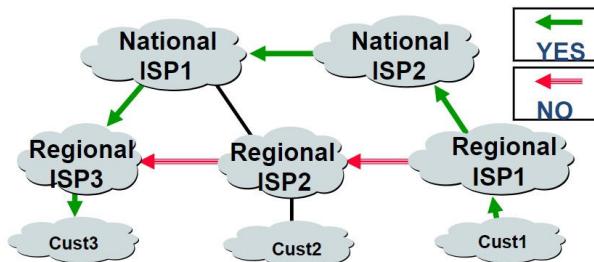
- Globální směrovací informace.
- Na síť je pohlíženo jako na množinu autonomních systémů.
- Používá se pro směrování mezi autonomními systémy.
- *Multi-metric.*

## Path Vector algoritmus

- Posílá celou cestu (posloupnost uzelů – autonomních systémů), ne jenom vzdálenost.
- Slouží také pro detekci smyček v rámci cesty.
- Cíl je projít přes co nejméně autonomních systémů.
- Umožňuje tzv. *flexible policies*, můžu se rozhodnout jakou cestou budu provoz směrovat. Část provozu můžu přes  $AS_1$ , jinou část pres  $AS_2$ , apod.
- Známe celou cestu směrování, víme přes co pakety půjdou, můžeme některý uzel vyřadit.
- Směrovací protokol Border Gateway Protocol (BGP).

- Route contains several different independent metrics

- Preferred exit from ASN
- Preferred entry to ASN
- ASN Path
- Origin
- Community



Obrázek 20.7: Path vector příklad.

# Kapitola 21

## PDS – Řízení toku dat (flow-control) a prevence zahlcení (congestion-control) na transportní vrstvě (MP-TCP, QUIC, SCTP, DCCP).

### 21.1 Zdroje

- 02-transportni-protokoly.pdf
- PDS\_2021-02-19.mp4

### 21.2 Úvod a kontext

**Transportní vrstva** Transportní vrstva je název čtvrté vrstvy modelu vrstvové síťové architektury (ISO/OSI model). Leží mezi vrstvou síťovou (L3) a aplikační (L7).

- Činnost na straně odesílatele: obdrží data z aplikační vrstvy, nasegmentuje je a ke každému segmentu/datagramu přidá L4 hlavičku.
- Činnost na straně příjemce: obdrží segmenty/datagramy, které uspořádá do správného pořadí (*out-of-order* doručení), předá je aplikační vrstvě.
- Komunikace mezi vrstvami L4 a L7 probíhá pomocí socketů.
- Zodpovědnost za *end-to-end* spojení.
- Adresuje aplikace pomocí portových čísel.
- *Quality of Service* – zotavení se z chyb, spolehlivost, řízení toku, řízení zahlcení.

**Multi homing** – Využítí více bodů připojení zároveň (pro jednu komunikaci). Pokud je podporováno a změní se IP adresa, nemusí se navázat nové spojení.

**Connection-oriented** – Typ spojení, kdy je nejprve nutné, navázat komunikaci. Týká se tří fáz: navázání spojení, přenos dat, ukončení spojení.

**Connection-less** – Typ spojení, které nerozlišuje jestli spojení existuje, nebo nikoliv. Pouze fáze přenosu dat.

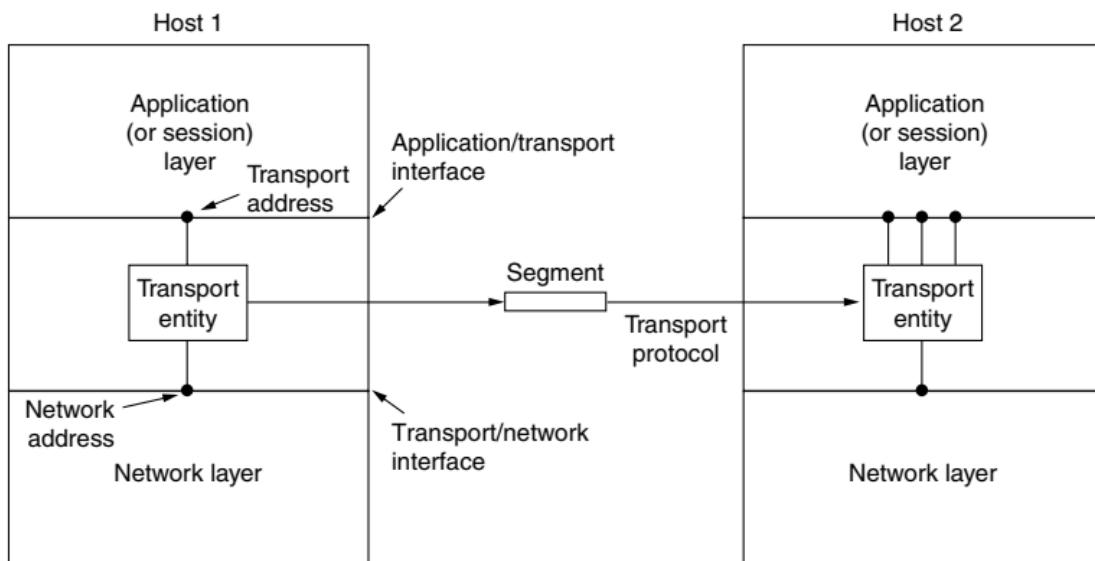
**Round trip time (RTT)** RTT je doba, která uplyne od vyslání paketu z jednoho komunikujícího uzlu k druhému po návrat potvrzení zpátky na první uzel.

### bandwidth, throughput, goodput

- *Bandwidth* – šířka pásma (maximální přenos dat danou cestou v jeden moment)
- *Throughput* – propustnost komunikace na úrovni protokolu
- *Goodput* – propustnost komunikace na úrovni aplikace (throughput – režie)

### Bitové chyby

- „Přeskočení bitu“ (*bit swapping*), jde o chyby na úrovni zpracování informace uzlem v síti (typicky spojené s přenosovým médiem)
- Zabýváme se jimi na úrovni L2
- Řešením je obvykle redundance (Hammingův kód, CRC, Viterbiho algoritmus, ...)



Obrázek 21.1: Transportní vrstva (L4).

### 21.3 Paketové chyby

- Jde o chyby na úrovni přenosu informace mezi dvěma uzly.
- Zabýváme se jimi na transportní vrstvě.
- Řešením je obvykle znovuzaslání paketu.
- Můžeme měřit:

- PER (*packet error rate*)

$$PER = \frac{\text{počet chybnych paketu}}{\text{počet prenesených paketu}}$$

- BER (*bit error rate*)

$$BER = \frac{\text{počet chybnych bitu}}{\text{počet prenesených bitu}}$$

- Dále jsou představeny druhy paketových chyb.

**Ztráta (zpoždění) paketu** Jak může dojít ke ztrátě (zpoždění) paketu?

- Neopravitelná bitová chyba (rámcem je zahozen na L2)
- Zahlcení linky (směrovače jsou přetíženy a některé pakety zahazují)
- Špatné směrovací tabulky (paket jde špatnou cestou, nebo cyklí a vyprší mu TTL)

**Ztráta fragmentovaných dat** Fragment se může ztratit ze stejného důvodu jako standardní nefragmentovaný paket.

**Duplicita paketu** Příjemce obdržuje stále paket se stejným sekvenčním číslem. Odesílatel si myslí, že se paket k příjemci nikdy nedostal. Proč?

- Ztratí se potvrzení o přijetí paketu.

**Vložení paketu** Příjemce obdrží paket, který do spojení nepatří. Jak k tomu může dojít?

- Přijetí zpožděného paketu, který dorazil až po skončení jednoho a začátku dalšího separátního datového toku.
- Podvrhávání paketu útočníkem snažícím se narušit integritu sítě.
- Paket je chybou „zvláštně zmrzačen“ tak, že chybu nelze rozpoznat (např. je změněm bit v cílové IP adrese) a je směrován jinému příjemci.

**Změna pořadí** Vychází z designu počítačových sítí, kdy paketu mohou proudit různými cestami a s různým zpožděním. Data jsou *out-of-order* a příjemce je musí přeuspřádat.

## 21.4 Řízení toku

Řízení toku (*flow control*) je řízení komunikaci mezi dvěma uzly na straně koncového systému. Řízení zahlcení (*congestion control*) je řízení komunikace v rámci sítě. Alternativně, se za řízení toku považuje koncept detekce a korekce paketových chyb. Často jsou tyto termíny vykládány různě.

### 21.4.1 Detekce paketové chyby

Paketové chyby detekujeme pomocí sekvenčních čísel. Sekvenční číslo je jedinečný identifikátor paketu v rámci datového toku, který identifikuje jeho pořadí. Tímto poznáme ztrátu, duplicitu i změnu pořadí.

Jak rozpoznat ztrátu od zpoždění?

- **Timeout** – Může být fixní, ale v lepším případě se odvozuje od RTT.
- **Negativní potvrzování** – Paket jsem dostal (ACK) vs. paket jsem nedostal (NACK). Pokud je přičina zpoždění zahlcení, tak tento přístup linku ještě více zahltí.

### 21.4.2 Korekce paketových chyb

Pokud paket nedorazil, odesílatel ho pošle znovu. Tzv. *Automatic Repeat/Request* (ARQ) – Čeká se na ACK, když nepřijde do timeoutu, dojde k znovuzaslání. Princip tzv. klouzavého okna (*sliding window*). Existují 3 strategie.

## Stop and wait

- Klouzavé okno o velikosti 1.
- Odesílatel pošle paket a čeká na potvrzení. Po přijetí potvrzení pošle další paket.
- Špatná efektivita využití pásma.

## Go back n

- Buffer na straně odesílatele.
- Příjemce potvrzuje naposledy přijatým paketem (např. příjemce pošle ACK2 znamená, že dostal pakety 0, 1, 2).
- Plýtvání při znovuzaslání (např. odesílatel pošle 5 paketů, ztratí se první, musí poslat znova všech 5).

## Selective repeat

- Buffery jsou na obou stranách.
- Efektivní využití šírky pásma, ale složitější implementace.
- Příjemce potvrzuje:
  - Fast Retransmit – Příjemce posílá potvrzení s naposledy přijatým paketem.
  - Bitová maska – V ACK je zavedeno další políčko (bitová maska), které obsahuje informace o tom, co příjemce přijal a co ne.
  - NACK – Příjemce posílá ACK, pokud balík dostal a NACK, pokud nikoliv.

## 21.5 Řízení zahlcení

Zahlcení je detekováno, co se s tím dá dělat?

- Zvýšit kapacitu sítě,
- Snížit množství provozu.

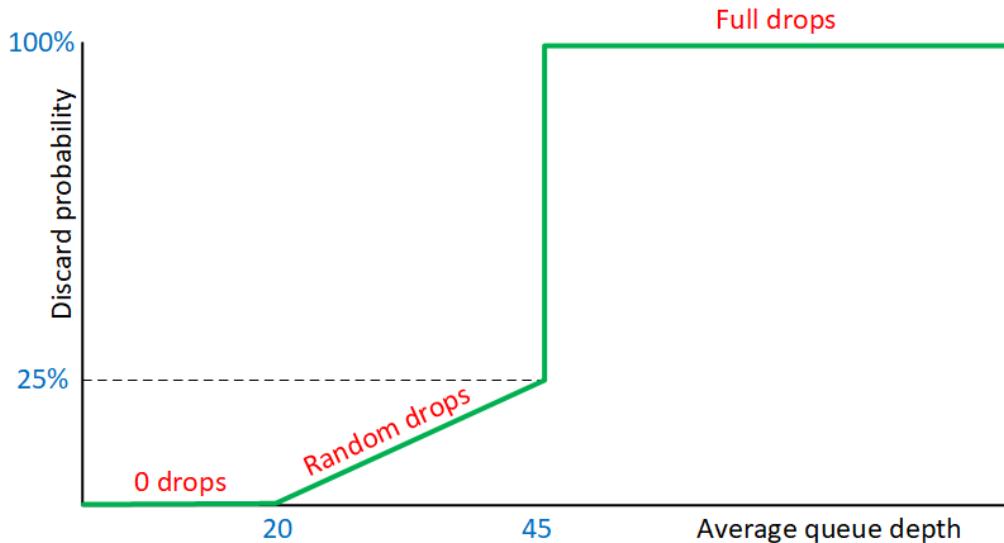
Jak se to dá dělat?

- Opravit zahlcený provoz (*repair*) – Je detekováno zahlcení a řešíme, co budeme dělat.
- Předcházet problému zahlcení (*avoidance*) – Snažíme se zahlcení předcházet (prevention).

### 21.5.1 Zahazování paketů (opravování provozu)

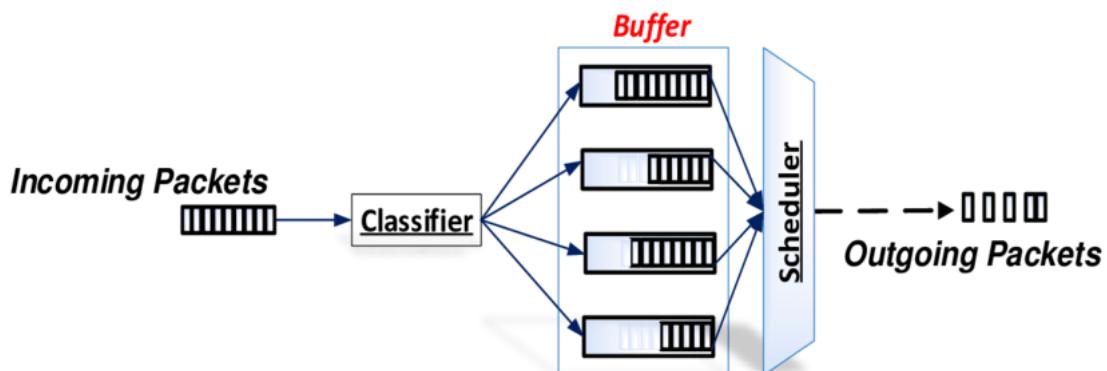
Každý uzel (nejčastěji směrovač) v síti kontroluje délku své fronty, pokud příliš narůstá, tak začne preventivně některé pakety zahazovat (to však vede k jejich znovuzaslání).

**RED (Random Early Detection)** Algoritmus RED sleduje velikost fronty a zahazuje pakety na základě pravděpodobnosti. Pokud je fronta téměř prázdná, pravděpodobnost zahodení je 0. Jak fronta roste, roste i pravděpodobnost zahodení příchozího paketu. Když je fronta plná, pravděpodobnost dosáhne hodnoty 1 a všechny příchozí pakety jsou zahazovány.



Obrázek 21.2: Vizualizace RED.

**WRED (Weighted Random Early Detection)** Algoritmus WRED je rozšířením RED o přidání váh pro jednotlivé třídy paketů (máme frontu pro každou třídu). WRED nezahazuje všechny pakety se stejnou pravděpodobností, ale rozlišuje jejich důležitost (pomocí hodnot IP precedence nebo DSCP). Některé pakety chceme prioritizovat, např. VoIP (*Voice over Internet Protocol*), RTP (*Real-time Transport Protocol*).



Obrázek 21.3: Vizualizace WRED.

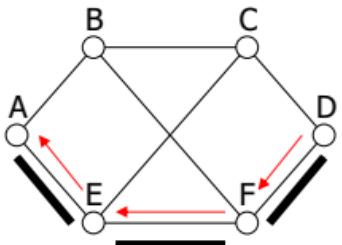
### 21.5.2 Choke pakety (opravování provozu)

Příjemce v síti má problém zpracovávat provoz (nestihá), tak dá vědět odesílateli (odesle choke paket), aby zpomalil.

**Plain choke packets** Viz obrázek 21.4. Uzel D detekuje zahlcení způsobené provozem od A. Uzel D pošle choke paket A. Proti zahlcení „zabaruje“ pouze odesílatel (A).

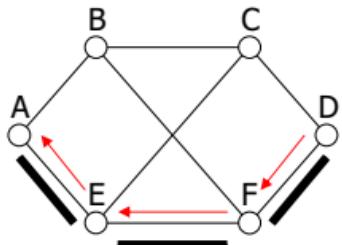
**Hop-by-hop choke packets** Viz obrázek 21.4. Uzel D detekuje zahlcení způsobené provozem od A. Uzel D pošle choke paket A. Každý uzel po cestě „bojuje“ proti zahlcení (F, E, A).

### Plain Choke packets



A heavy flow is established  
Congestion is noticed at D  
A Choke packet is sent to A  
The flow is reduced at A  
The flow is reduced at D

### Hop-by-hop Choke packets



A heavy flow is established  
Congestion is noticed at D  
A Choke packet is sent to A  
The flow is reduced at F  
The flow is reduced at D

Obrázek 21.4: Plain choke packets vs. hop-by-hop choke packets.

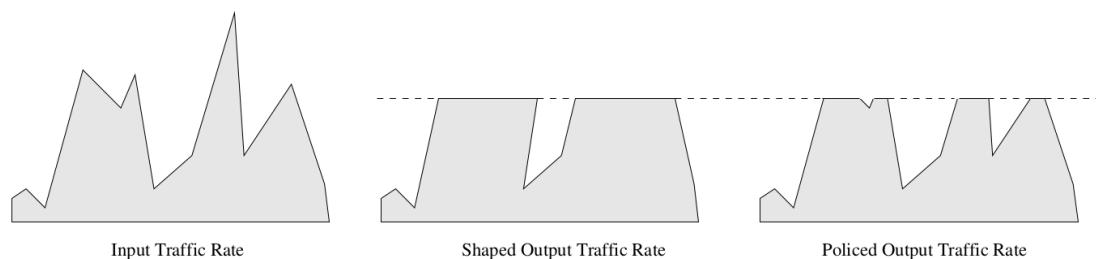
### 21.5.3 Ořezání a rozložení provozu (předcházení zahlcení)

Jedná se o předcházení zahlcení ze strany odesílatele. Odesílatel je „ohleduplný“ a myslí na to, aby příjemce nebyl zahlcen. Odesílatel má pevně stanovenou hranici kolik může maximálně posílat. Implementace pomocí tzv. *token bucket*.

**Ořezání provozu (policing)** Cokoliv nad hranici je „oríznuto“ a zahozeno.

**Rozložení provozu (shaping)** Cokoliv nad hranici je uloženo do bufferu a odesláno později, až je menší provoz.

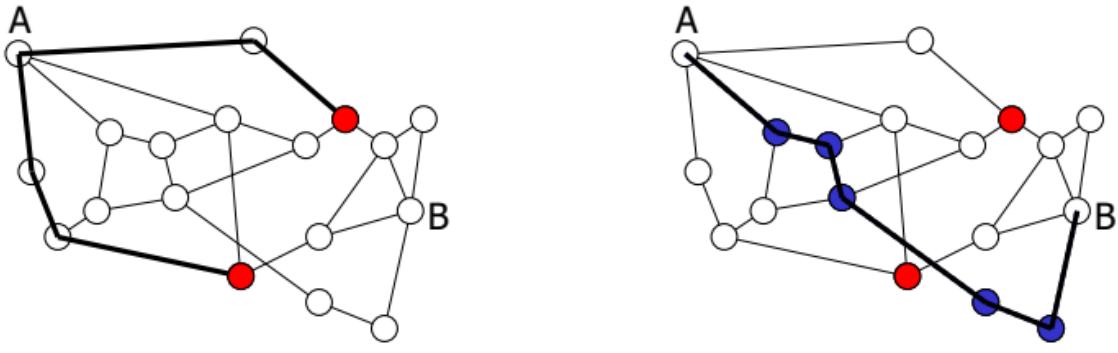
Čím se liší rozložení provozu (policing) a omezení provozu (shaping)?



Obrázek 21.5: Ořezání provozu vs rozložení provozu.

### 21.5.4 Rezervace (předcházení zahlcení)

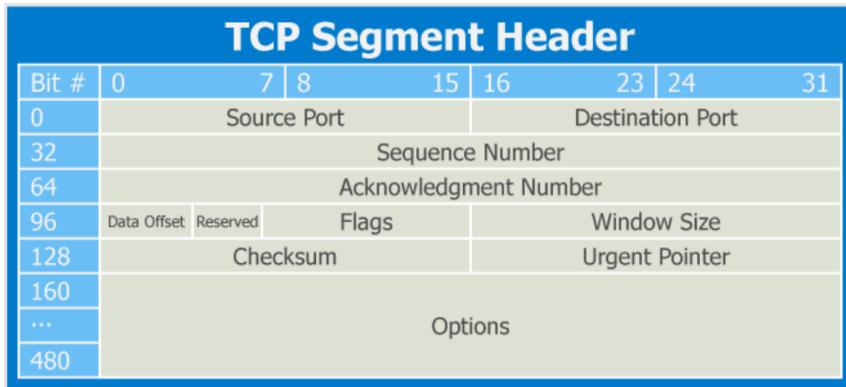
Zařízení v síti (směrovače) jsou schopni se domluvit a vytvořit si tzv. „virtuální okruh“ pro konkrétní komunikaci. Každé zařízení na okruhu ví o provozu a rezervuje potřebné pásmo (předem se deklarují kapacity). Je nutné více signaliace (režie).



Obrázek 21.6: Příklad rezervace.

## 21.6 TCP (*Transmission Control Protocol*)

- Garantuje spolehlivé doručení.
- Garantuje doručení v pořadí.
- Řízení toku a zahlcení.
- *Connection-oriented* – Navázání spojení pomocí *three-way handshake*.
- Pracuje s *byte stream* – nezohledňuje hranice aplikačních dat.
- Sekvenční číslo závisí na tom, kolik bajtů se posílá. Např.  $seq = 92, data = 8 B, seq = 100, data = 20 B, seq = 120, data = 13 B, \dots$

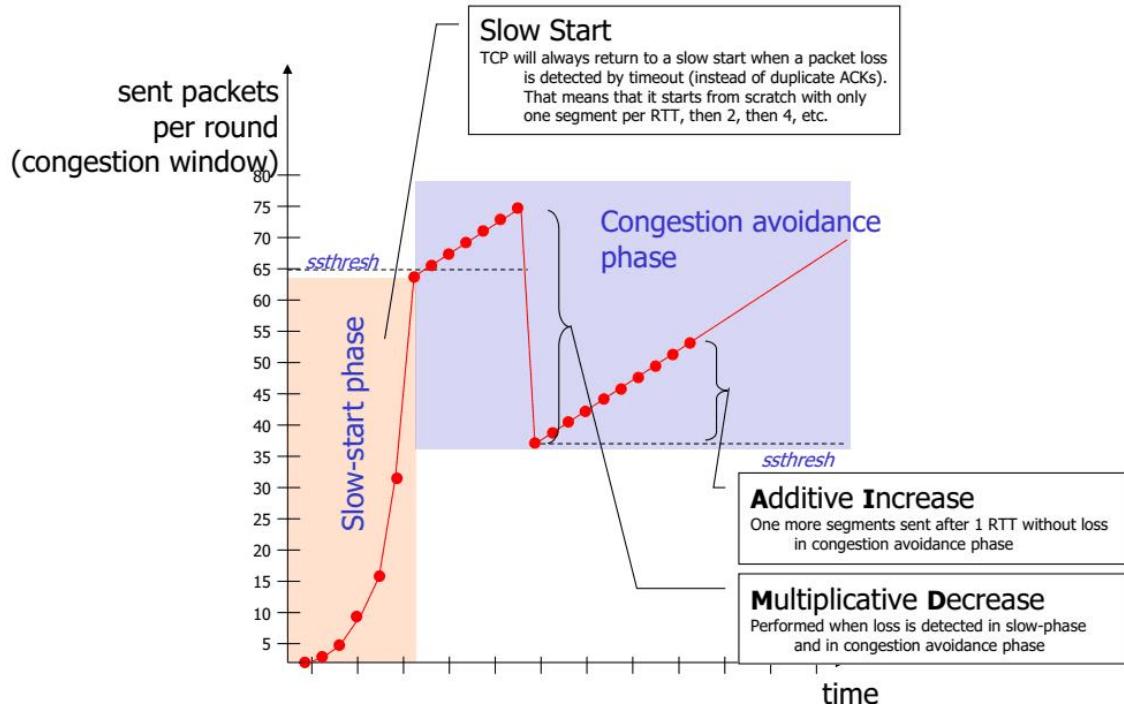


Obrázek 21.7: Hlavice TCP.

**Řízení toku a zahlcení** V TCP jsou rozlišovány 3 fáze přenosu dat.

- *Slow start* – Exponenciální zvyšování rychlosti odesílání až po dosažení nějaké hranice.
- *Congestion avoidance (additive increase)* – Lineární zvyšování počtu odeslaných paketů.
- *Congestion avoidance (multiplicative decrease)* – Skokové snížení rychlosti odesílání. Nastane, pokud je detekována ztráta paketu.

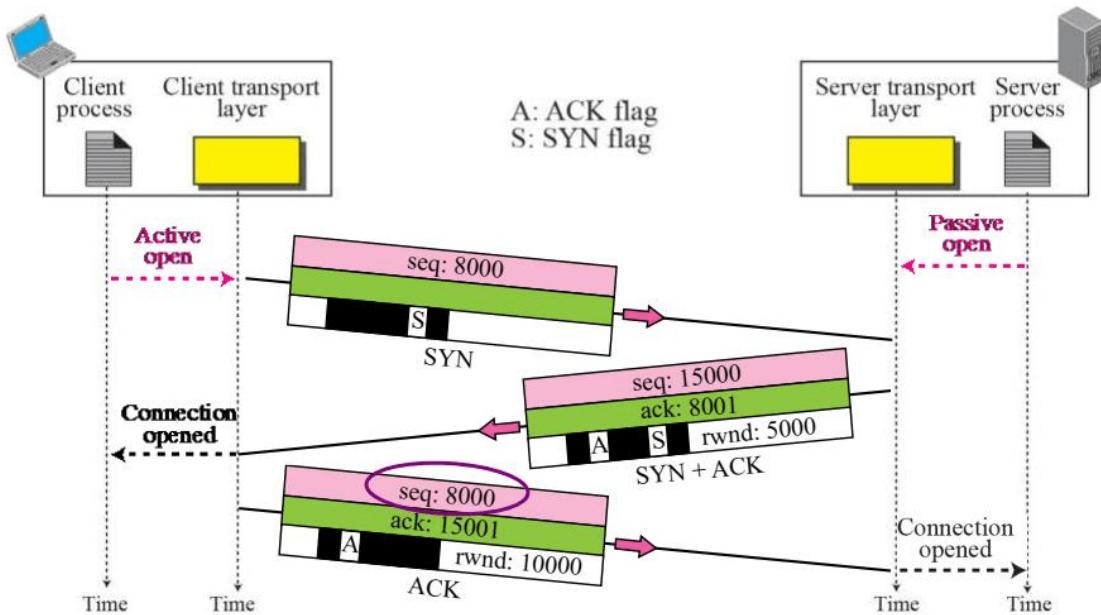
V TCP existuje spoustu algoritmů pro řízení toku a zahlcení, které implementují tento princip (Tahoe, Reno, New Reno, Vegas, CUBIC, Westwood, ...).



Obrázek 21.8: Jednotlivé fáze řízení zahlcení v TCP.

### Nevýhody a problémy

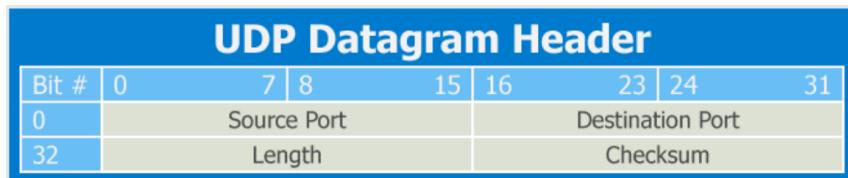
- Nepodporuje *multi homing*.
- *Head of line blocking* – Pokud dojde ke ztrátě paketu, tak vše za ním je pozdrženo
- Velká režie potvrzování (až 35 % veškerého provozu je režie)



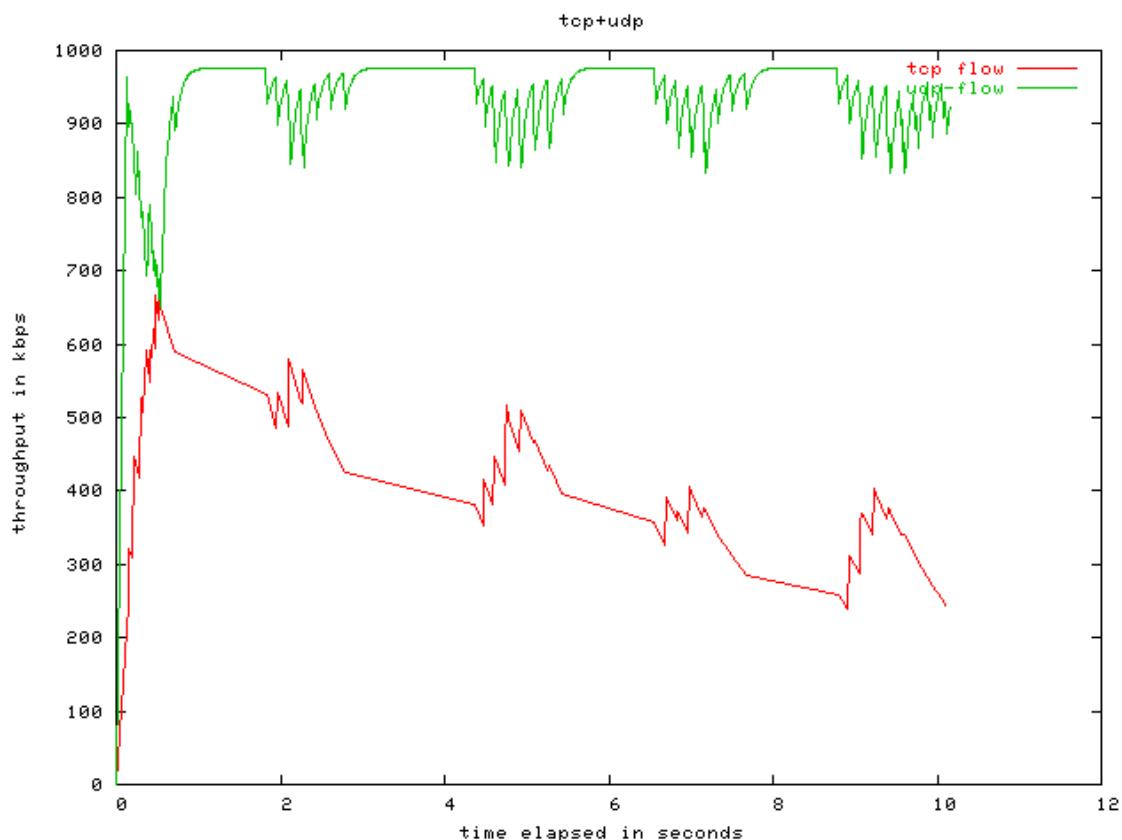
Obrázek 21.9: Zahájení spojení pomocí tzv. *Three-Way Handshake*.

## 21.7 UDP (*User Datagram Protocol*)

- Negarantuje spolehlivé doručení (*best effort*).
- Negarantuje doručení v pořadí.
- *Connection-less*.
- Pracuje s *byte stream* – nezohledňuje hranice aplikačních dat.
- Má minimální režii.



Obrázek 21.10: Hlavička UDP.

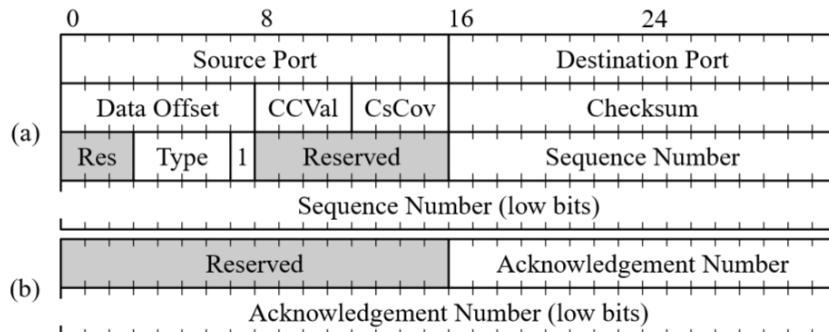


Obrázek 21.11: Problém UDP, kdy si pro sebe zabere celé pásmo, protože TCP se chová „zodpovědně“ a snižuje svůj provoz.

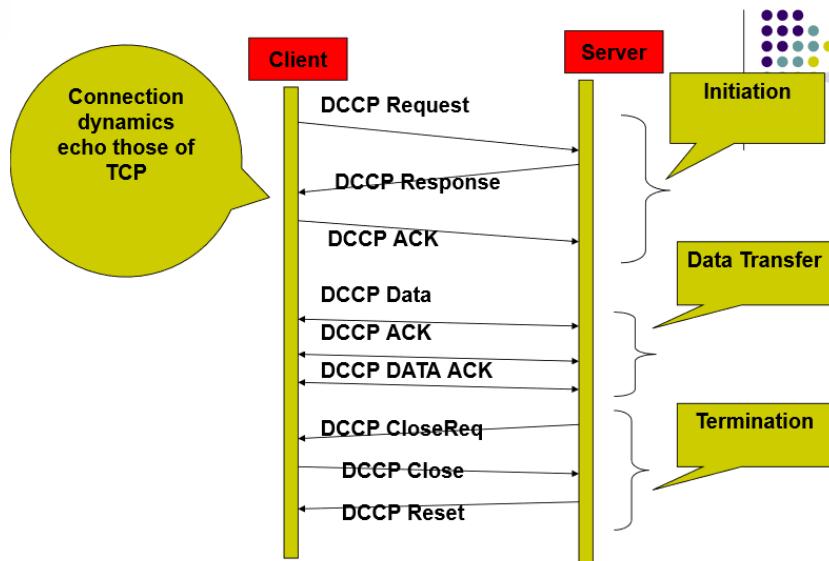
## 21.8 DCCP (Datagram Congestion Control Protocol)

- Cílem je zajistit podporu řízení zahlcení pro UDP.
- Negarantuje spolehlivé doručení (*best effort*).

- Negarantuje doručení v pořadí.
- *Connection-oriented* – Navázání spojení pomocí *three-way handshake*.
- Využití pro audio/video konference.



Obrázek 21.12: Hlavička DCCP. Sekvenční a potvrzovací čísla souvisí pouze s účtováním provozu kvůli řízení zahlcení, nikoliv pro zajištění spolehlivého přenosu.



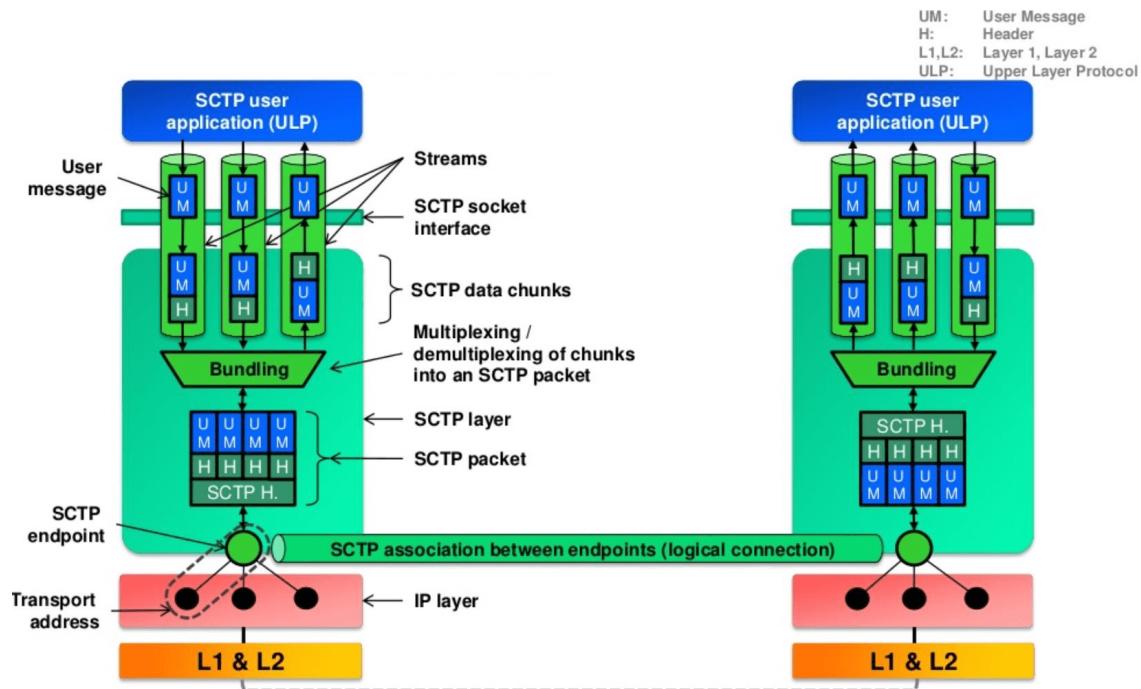
Obrázek 21.13: Komunikace pomocí DCCP. Obsahuje několik typů zpráv: Request, Response, ACK, Data, Data ACK, CloseReq, Close, Reset.

**Jak je zahlcení realizováno** V potvrzeních se nastavují tzv. ECN (*Explicit Congestion Notification*) bity. Pomocí nich dává příjemce odesílateli vědět, jak moc je zahlcen. V průběhu navazování spojení se pomocí CCID (*Congestion Control ID*) dohodne, jaký typ řízení zahlcení se bude používat (algoritmy jako pro TCP, nebo vlastní).

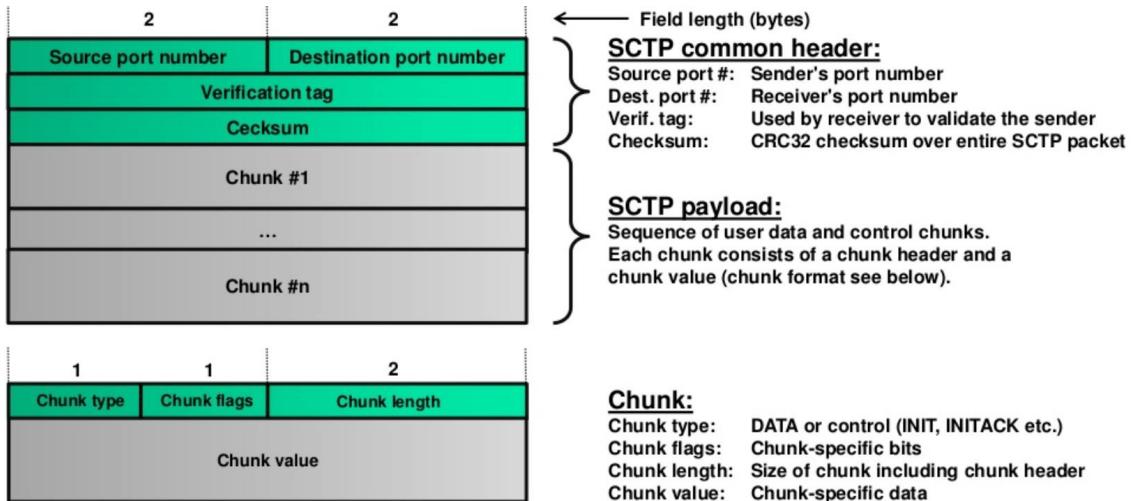
## 21.9 SCTP (*Stream Control Transmission Protocol*)

- Garantuje spolehlivé doručení.
- Garantuje doručení v pořadí.

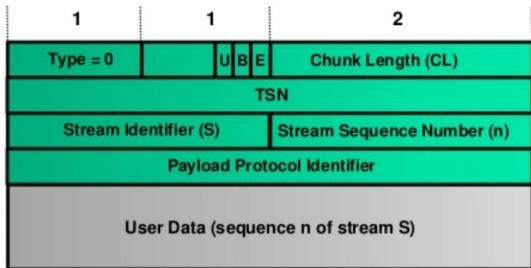
- Pracuje s *message stream* – Respektuje hranice aplikačních dat.
- *Connection-oriented* – Navázání spojení pomocí *four-way handshake*.
- *Path MTU discovery* – Obrana proti fragmentaci (od zdroje k cíli jsou prozkoumány MTU všech uzlů).
- Netrpí problémem *head of line blocking*.
- Podporuje *multi homing*.
- Řízení zahlcení.
- Aplikace je třeba přeprogramovat, aby místo TCP socketů využívali SCTP sockety.



Obrázek 21.14: Komunikace pomocí SCTP. Komunikace s aplikací probíhá přes sockety. Každé aplikační zprávě je přidán základní hlavičkou – vzniká *data chunk*. Několik data chunku je zabaleno dohromady a opatřeno SCTP hlavičkou – vzniká SCTP paket. Ten je poté možno posílat přes více IP spojení.



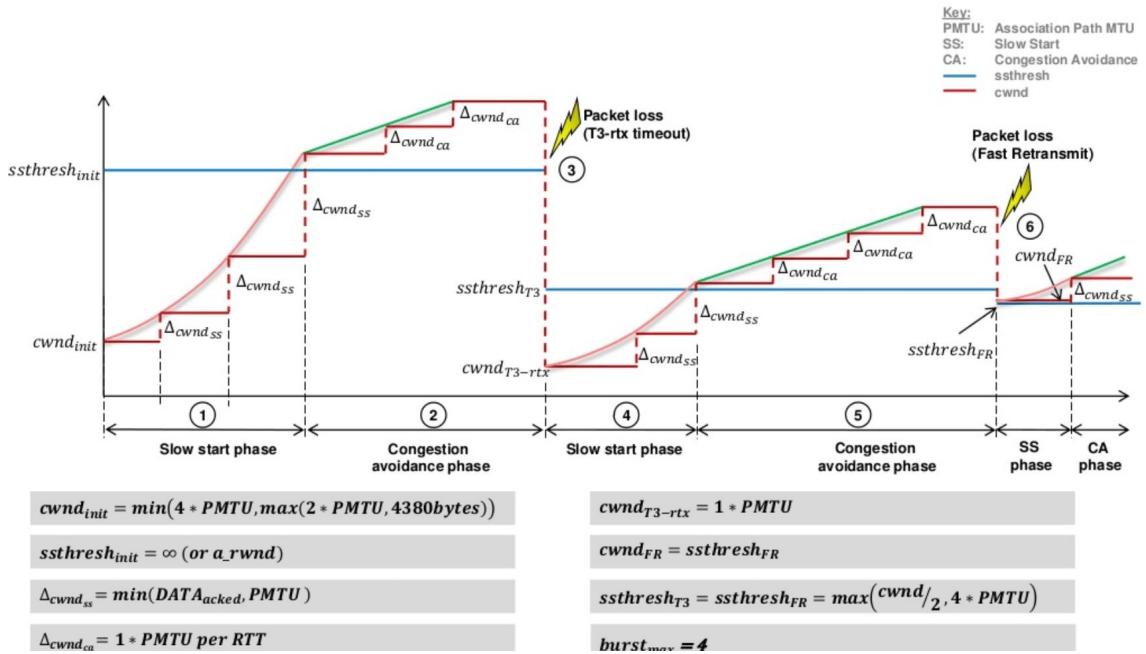
Obrázek 21.15: Hlavička SCTP. SCTP paket a *data chunk*.



#### Data chunk:

- U bit: If set to 1, indicates that this is an unordered chunk. Unordered chunks are transmitted and sent to the receiving application as is without re-ordering based on the sequence number.
- B bit: Begin of fragment bit. If set to 1 this is the first fragment of a larger fragmented user data message.
- E bit: End of fragment bit. If set to 1 this is the last fragment of a larger fragmented user data message.
- Stream identifier: Identifies the stream to which this chunk belongs.
- Stream seq. no.: Sequence number of user data within this stream. In case of fragmentation this numer is identical for all fragments.
- Payload proto. id.: Identifies the upper (application) layer protocol (e.g. HTTP).
- User data: Application user data (e.g. HTTP header and payload).

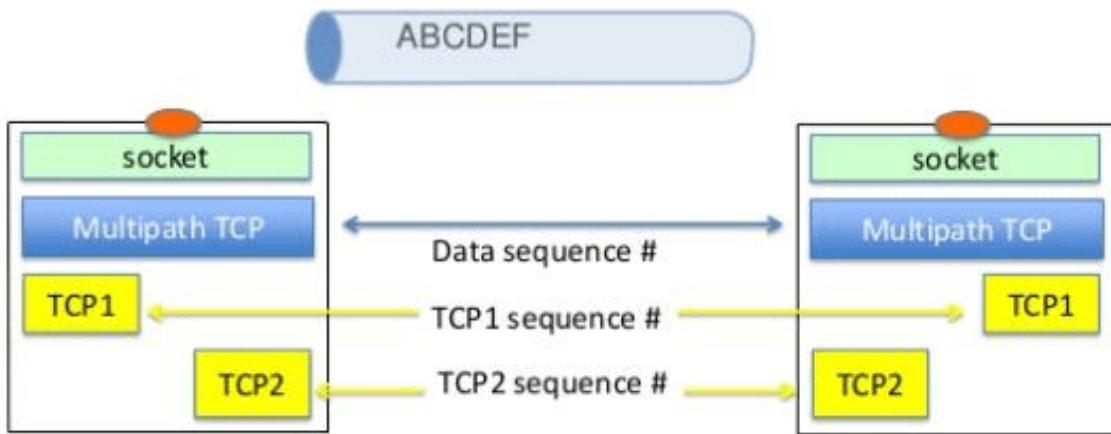
Obrázek 21.16: Detailní hlavička *data chunk*. Identifikátor streamu a identifikátor paketu v rámci streamu.



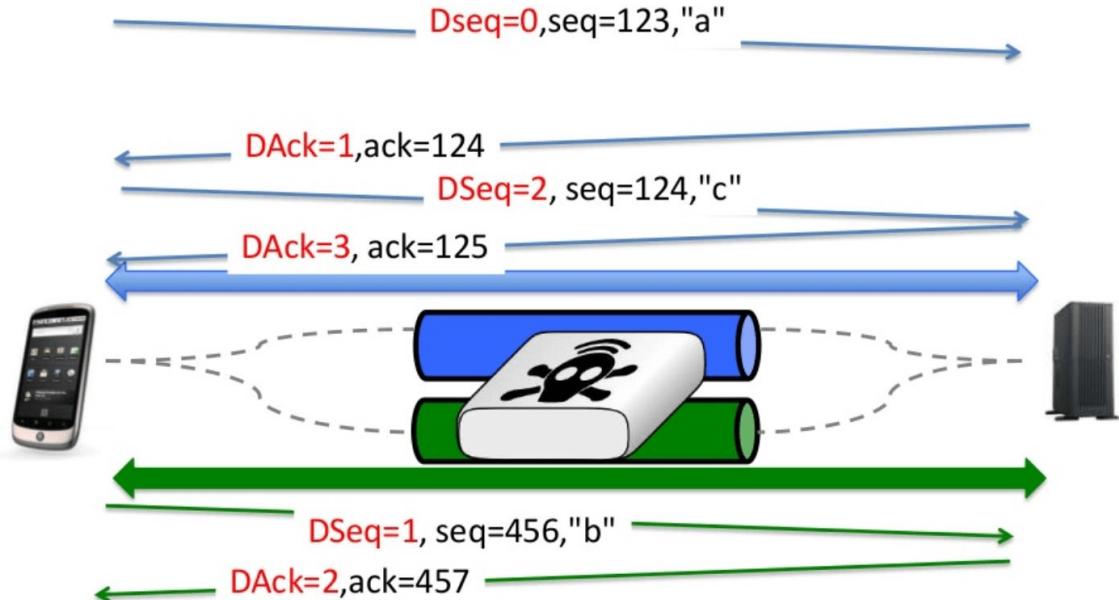
Obrázek 21.17: Řízení zahlcení (*slow start*, *congestion avoidance (additive increase)*, *congestion avoidance (multiplicative decrease)*) je podobné jako TCP Westwood.

## 21.10 MP-TCP (*Multipath TCP*)

- Cílem je vzít co nejvíce vlastností z SCTP a poskytnout je v TCP bez toho, aniž by se musely přeprogramovávat aplikace (má obdobné sockety jako TCP).
- Má menší signalizační náročnost oproti SCTP.
- Chová se jako TCP pouze, podporuje *multi homing* a má vlastní řízení zahlcení.
- Rozšíření TCP hlavičky o tzv. *options*.



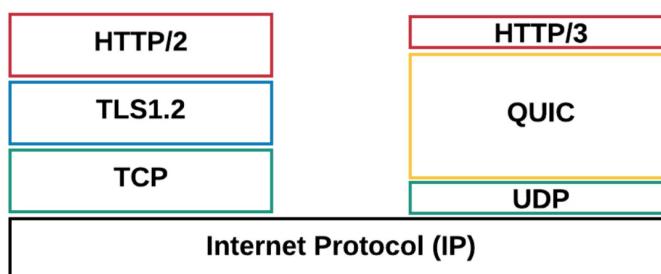
Obrázek 21.18: MPTCP komunikace, rozdělení aplikacních dat na paralelní TCP streamy. Rozlišujeme datová sekvenční čísla a sekvenční čísla TCP.



Obrázek 21.19: MPTCP řízení toku. Dvě TCP komunikace a posílání řetězce „abc“. Navázání spojení podobně jako v TCP. Může vypadnout i celý komunikační kanál.

## 21.11 QUIC (*Quick UDP Internet Connections*)

- Motivace pro QUIC: dnes většina provozu probíhá šifrované přes HTTPS. Navázání bezpečného HTTPS spojení má velkou režii (HTTPS = TCP + SSL + HTTP). Až  $3 \times \text{RTT}$  přes započetím komunikace!
  - TCP: syn, syn ack, ack
  - TLS: client hello, server hello, change cipher spec, encrypted handshake message
- Staví nad UDP (aby všechny zařízení v internetu nemuseli implementovat nový transportní protokol).
- Paralelní datové streamy, aby bylo zabráněno *head of line blocking*.
- Má zabudovanou *Forward Error Correction* (FEC), přidává bitovou redundanci, tím umožňuje velké množství ztracených paketů dopočítat. Pokud to není možné, tak „chytrý“ selective repeat.



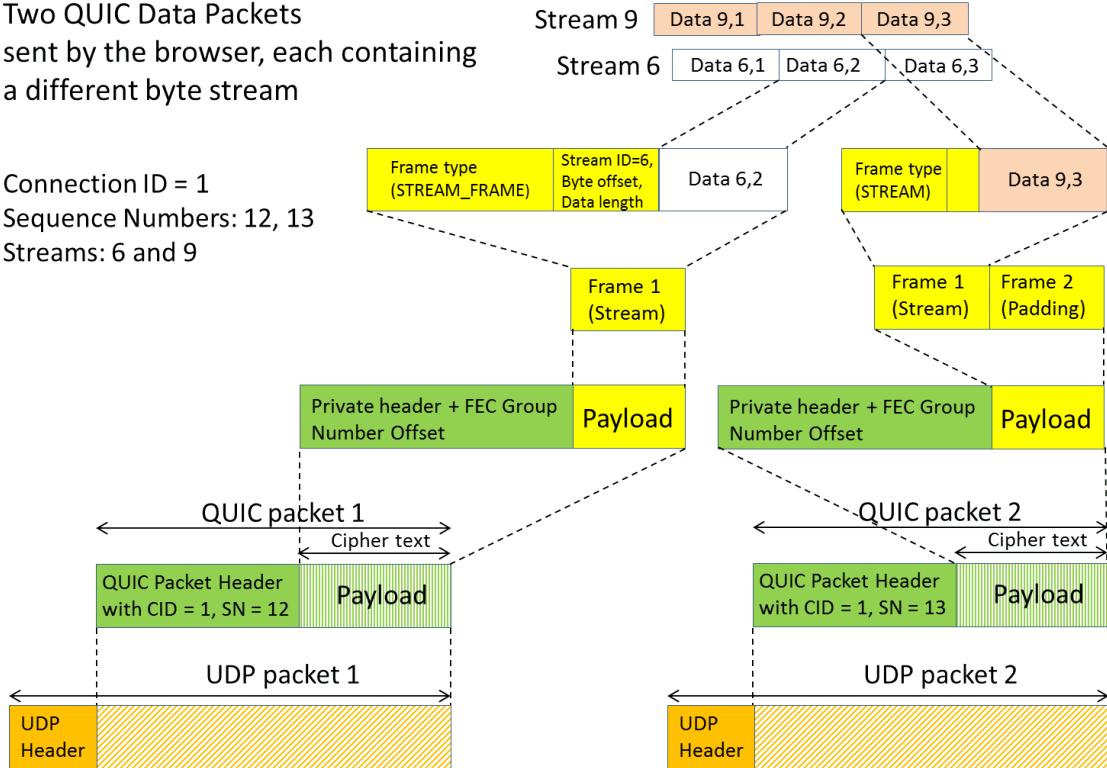
Obrázek 21.20: Standardní TCP stack vs. QUIC.

|                    |   |   |   |   |   |   |   |   |   |                                 |   |   |   |   |   |   |   |   |   |   |   |
|--------------------|---|---|---|---|---|---|---|---|---|---------------------------------|---|---|---|---|---|---|---|---|---|---|---|
| 0                  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0                               | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 |
| public flag (1)    |   |   |   |   |   |   |   |   |   | connection id (0,1,4,8)         |   |   |   |   |   |   |   |   |   |   |   |
| version(0,1)       |   |   |   |   |   |   |   |   |   | sequence num(1,4,6,8)           |   |   |   |   |   |   |   |   |   |   |   |
| frame_type(0x80~)  |   |   |   |   |   |   |   |   |   | private flag(1) FEC offset(0,1) |   |   |   |   |   |   |   |   |   |   |   |
| stream_id(1,2,3,4) |   |   |   |   |   |   |   |   |   | offset(0, 2,3,4,5,6,7,8)        |   |   |   |   |   |   |   |   |   |   |   |
| data length(0, 2)  |   |   |   |   |   |   |   |   |   | Stream DATA                     |   |   |   |   |   |   |   |   |   |   |   |

Obrázek 21.21: Hlavička QUIC. Spousta políček má proměnlivou velikost.

Two QUIC Data Packets  
sent by the browser, each containing  
a different byte stream

Connection ID = 1  
Sequence Numbers: 12, 13  
Streams: 6 and 9



Obrázek 21.22: Složení QUIC zprávy.

## Kapitola 22

# PDS – Metody detekce sít'ových incidentů (signatury, statistické metody) a nástroje (IDS/IPS).

### 22.1 Zdroje

- 07-ids.pdf
- PDS\_2021-03-26.mp4
- 03-IDS.pdf
- NSB\_2021-02-23.mp4

### 22.2 Úvod a kontext

#### Identifikace a monitorování sít'ového provozu

- Co je cílem?
  - Identifikace protokolu (na různých vrstvách).
  - Identifikace aplikace.
  - Pokud je identifikován nedovolený provoz, tak kontaktovat správce.
- Proč to dělat?
  - Bezpečnost – Detekce útoků a nedovoleného chování.
  - Diagnostika – Správné chování systémových i uživatelských aplikací.
  - Rozlišení služeb – Nastavení kvality služeb, prioritizace kritických přenosů.
  - Vytížení sítě – Sledování rozložení sítě, vytížení sít'ových prvků a služeb.
  - Účtování služeb – Identifikace provozu VoIP, IPTV, apod.

**Sít'ové tunelování** Technika, která pro přenos sít'ového spojení používá jiné sít'ové spojení (Zapouzdření komunikace do jiného protokolu.). Umožňuje: přenášet data přes nekompatibilní sítě; obcházet administrativní omezení určité sítě; poskytovat zabezpečenou komunikaci přes nezabezpečenou, resp. nedůvěryhodnou sít'.

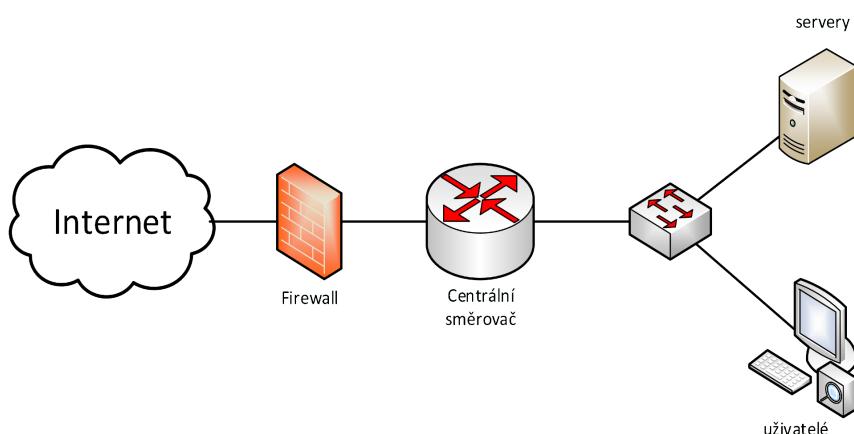
## 22.3 Nástroje IDS/IPS

Umištěno někde v síti, aby se zabezpečili všechny zařízení najednou. Specializované zařízení, např. IDS/IPS sonda nebo jako součást hraničního prvku.

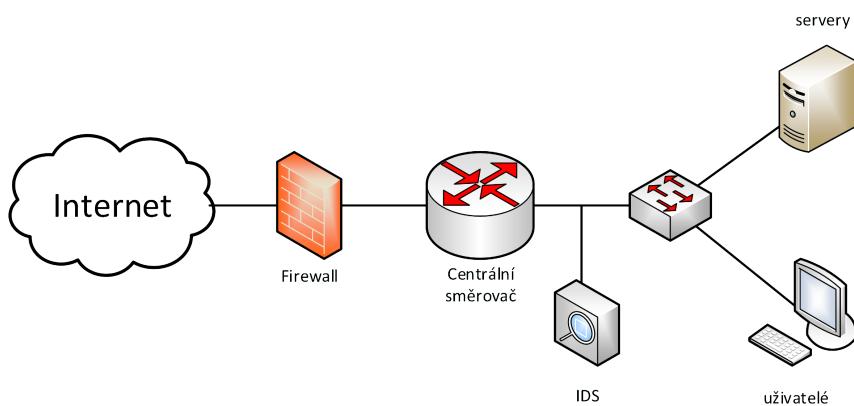
**Intrusion Detection System (IDS)** „Pasivní obraný mechanismus“. Cílem je identifikovat provoz a v případě, že se jedná nedovolený provoz, tak informovat správce. Ten rozhodne, jaké budou kroky (zda se zablokuje, nebo zda se jedná o falešnou pozitivitu).

**Intrusion Prevention System (IPS)** „Aktivní obraný mechanismus“. Cílem je detekovat provoz a v případě, že se jedná nedovolený provoz, tak ho odfiltrovat/zablokovat/zahodit. Nebezpečné v případě falešné pozitivity. Můžeme odříznout provoz, který považujeme za nedovolený, ale ve skutečnosti se jedná o provoz, který jsme nečekali. Většinou se v praxi volí IDS.

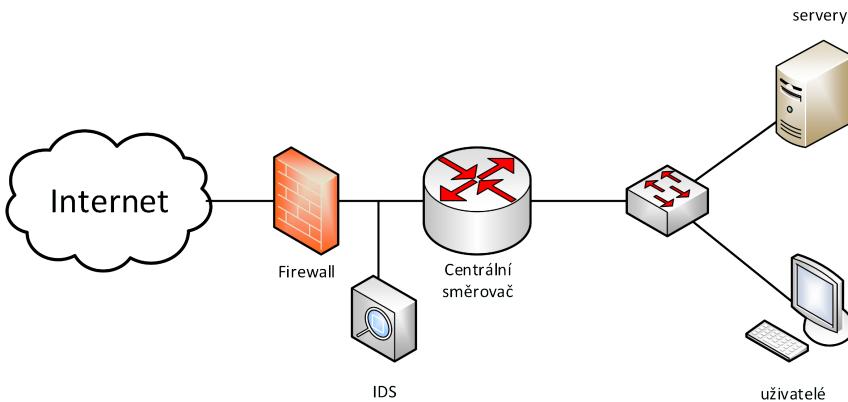
**Sítový incident** Identifikace nedovoleného provozu.



Obrázek 22.1: Způsob zapojení 0. Host based IDS/IPS. Systém nainstalovaný na koncové stanice a reportující informace. „Lepší firewall, lepší antivirus“.



Obrázek 22.2: Způsob zapojení do vnitřní sítě. Network based IDS/IPS.



Obrázek 22.3: Způsob zapojení před vnitřní sítí'. Network based IDS/IPS. Problém s NATem – nerozlišíme jestli komunikace jde na jeden uzel v síti nebo na více. Problém šifrování.

## 22.4 Identifikace provozu pomocí hodnot z hlaviček paketů

- Využití informací z hlaviček jednotlivých protokolů TCP/IP. Jaké informace lze využít?
  - Na linkové vrstvě (L2):
    - \* Field **EtherType** – IPv4, IPv6, ARP, RARP, autentizace 802.1x, ...
  - Na sít'ové vrstvě (L3):
    - \* Field **protocol** – IPv6 nad IPv4, ICMP, IGMP, transportní protokoly (UDP, TCP, SCTP, ...), směrovací protokoly (EIGRP, OSPF, ...), multi-cast (PIM), ...
  - Na transportní vrstvě (L4):
    - \* Field **SourcePort**, **DestinationPort**, rozpoznání služby na základě portu – FTP, SSH, Telnet, SMTP, DNS, DHCP, HTTP, POP3, NTP, IMAP, SNMP, LDAP, HTTPS, ...
  - Na aplikační vrstvě vrstvě (L7):
    - \* Specifické pro aplikace (např: pro HTTP rozpoznávání jednotlivých metod).
- Rychlé, dobře implementovatelné a univerzální řešení. Má své limity:
  - Aplikace nemusí být pevně svázána s portem (dynamické porty).
  - Tunelování, šifrování (spoustu věcí je zapouzdřováno do SSL/TLS), VPN.
  - IPv6 nad IPv4.
  - Skryté kanály – Tunelování skrze protokoly, kde se neočekává nic nelegitimního, běžně se nefiltrují (ICMP, DNS).
- Na tomto konceptu fungují firewally.

## 22.5 Identifikace provozu dle signatury

- Signatura (pattern, signature) je statický řetězec tvořený posloupností znaků z hla- vičky či těla protokolu, který slouží k rozpoznání daného protokolu.
  - Je to „reprezentující“ podřetězec paketu komunikace (paket jako řetězec bytů – ASCII).
  - Určujeme až v aplikačních datech (TCP/UDP payload).
- Signatura ale může mít různou podobu:
  - Řetězec, který se vyskytuje na začátku, resp. na nějakém fixním offsetu paylo- adu.
  - Řetězec, který se vyskytuje na různém místě v payloadu.
  - Posloupnost podřetězců, která se vyskytuje v payloadu.
- Pokud se snažíme identifikovat provoz na základě signatury, tak typicky signaturu nehledáme v celém paketu (příliš náročné), ale pouze v prvních X bytech payloadu.
- Signatura v paketu vs. toku
  - Jednoduchá možnost – Signatura pouze na základě informací v paketu.
  - Složitější možnost – Signatury napříč pakety (v datovém toku).
    - \* Příklad: signaturu pro identifikaci TLS provozu, by mohla mít podobu dvou řetězců, které se vyskytují každý v jiném paketu (1 – „client hello“, 2 – „server hello“).
- S novou verzí protokolu, je většinou třeba aktualizovat signaturu, protože se změnila struktura komunikace.
- Lze vytvořit manuálně na základě specifikace protokolu a nebo automatizovat (*Longest common subsequence problem*).
- Použití: Někde poslouchám síťový provoz a mám databázi signatur. Snažím se najít v toku signaturu z databáze a tím identifikovat aplikaci.

```
1 # First thing that happens is that the client sends NICK and USER, in
2 # either order. This allows MIRC color codes (\x02-\x0d instead of
3 # \x09-\x0d).
4 ^nick[\x09-\x0d -~]*user[\x09-\x0d -~]*:|user[\x09-\x0d -~]*:[\x02-\x0d -~]
5 *nick[\x09-\x0d -~]*\x0d\x0a)
```

Výpis 22.1: Příklad signatury pro detekci IRC. Pomocí regulárního výrazu je specifikována kostra řetězce.

```
1 ^(\x13bitTorrent protocol|azver\x01$|get /scrape\?info_hash=get
2 /announce\?info_hash=|get /client/bitcomet/|GET /data\?fid=)
3 |d1:id2:id20:|\x08'7P\[RP]
```

Výpis 22.2: Příklad signatury pro detekci BitTorrentu. Signatura je založená na klíčových slovech.

```
1 // SSL Hello with certificate, Client Hello  
2 ^(.?.?\x16\x03.*\x16\x03|.?.\x01\x03\x01?.*\x0b)
```

Výpis 22.3: Příklad signatury pro detekci SSL provozu. Detekce Client Hello paketu, kdy se vyměňují informace o šifrování.

### Automatizace vytváření signatur

- Algoritmus LCS (Longest Common Subsequence).
- Algoritmus hledá „nejbližší“ společný řetězec v toku (algoritmus definuje metriku blízkosti).
- Proces vytvoření signatury:
  - Odchytím si typickou komunikaci aplikace, kterou chci identifikovat v provozu – Několik toků, které jsou dostatečně reprezentativní.
  - Z každého toku vezmeme X prvotních paketů (např. 10 stačí).
  - Poté iterativně hledám „nejbližší“ společný podřetězec vždy pro dvojici toků (kandidát).
  - Cyklus končí, až se kandidát přestane měnit. Výsledek je signatura (minimální možná délka, 2 znaky jsou málo).

## 22.6 Identifikace provozu dle statistického chování

- Vytvoření statistického modelu protokolu na základě trénovacích dat.
  - Popis chování na základě metadat o provozu (odezva, velikost paketu, počet paketů v daném směru, ...).
  - Zkoumáme prvních  $n$  bytů toku, které jsou spíše specifické pro protokol (navazování spojení).
  - Na rozdíl od signatur se nezkoumá obsah hlaviček a těla protokolu, ale pouze statistické vlastnosti jeho chování.
- Klasifikace neznámého protokolu: Vybereme model s nejlepší vzdáleností od neznámého protokolu.
- Vytváření se tzv. **otisk protokolu** (*fingerprint*) – soubor vlastností (chování) protokolu, které lze využít k jednoznačné identifikaci protokolu.

### Metoda jednoduchých statistických otisků

- Trénovací data: mějme  $n$  síťových toků. Z prvních  $m$  paketů každého toku se extrahuje:
  - velikost paketu;
  - časový odstup paketů (závisí na tom co dělá klient a co dělá server);
  - pořadí paketů v toku.
- Učení:

- Z trénovacích dat se modeluje statistické chování každého paketu v toku. Např. velikost prvního paketu odpovídá tomuhle pravděpodobnostnímu rozdělení, časový odstup druhého paketu od prvního odpovídá jinému pravděpodobnostnímu rozdělení. (Typicky normální rozdělení).
- Poté se aplikují gaussovské filtry pro odstranění šumu.
- Vzniká tzv. „maska“ a je určena nějaká prahová hodnota (na základě *anomaly score*, která se spočítá z vektoru anomalií).
- Model: protokol je reprezentován otiskem (*fingerprint*) – „maska“, prahová hodnota
- Klasifikace:
  - Máme neznámý tok, který chceme klasifikovat.
  - Sleduju velikosti paketů, odstupy paketů a na základě toho spočítám *anomaly score* pro tok.
  - Získané *anomaly score* porovnám s prahovou hodnotou.

### Příklad:

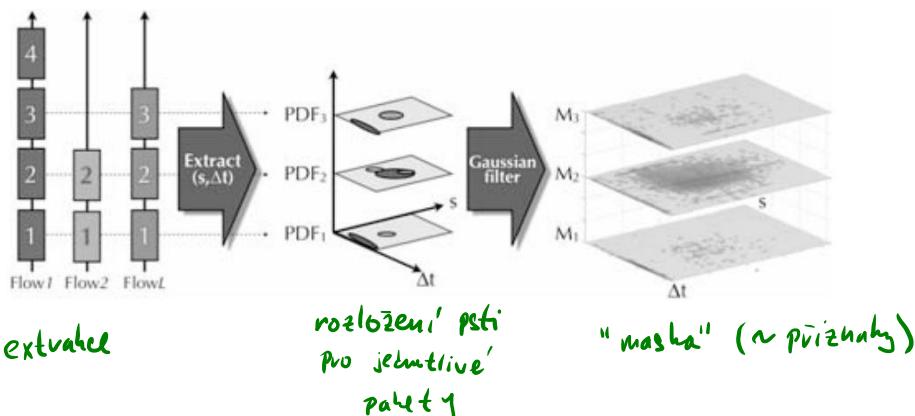
• Tok  $F = (P_1, P_2, \dots, P_n)$

• paket  $P_i = (s_i, \Delta t_i)$   
 $s_i$  - velikost paketu

$\Delta t_i$  - časový odstup od  $P_{i-1}$

• Je sestaven model hustoty pravděpodobnosti

• Trénování:



Obrázek 22.4: Princip metody jednoduchých statistických otisků – fáze učení.

### Statistical Protocol IDentification (SPID)

- Kombinuje statistické chování s atributy.
- Statistické atributy: směr paketů, pořadí paketů, velikost paketů.
- Aplikační atributy: četnost konkrétních bytů, hodnota off-setu.
- Model protokolu používá pro klasifikaci pravděpodobnostní vektory atributů.

- Pro porovnání měří odlišnost dvou pravděpodobnostních rozložení.
- Příklad jaké atributy přenosu lze použít pro rozlišení aplikačního protokolu.
  - Bytová frekvence prvního paketu TCP v každém směru (myšlenka: obsah je typický, pouze pro začátek toho toku – navázání spojení).
  - Počet změn komunikace (klient → server, server → klient).
  - Počet přenesených bytů v daném směru (RTP – balanced, HTTP – download, SMTP – upload).

## 22.7 Detekce anomálie

- Protokol jsme již identifikovali a nyní sledujeme jeho chování.
- Anomálie v provozu – odchylka komunikace od standardního (typického) provozu.
- Jaké anomálie mohou nastat:
  - Neočekávaný nárůst provozu.
  - Snížení provozu, chybějící linky – Výpadek linky.
  - Změna struktury provozu – Jeden uzel se snaží komunikovat se všemi v síti (sledování, detekce malwaru).
- Vytvoření modelu:
  - Trénovací data: typická komunikace.
  - Natrénování modelu.
  - Metrika pro určení odchylky od natrénovaného modelu (např. když se přesáhne práhová hodnota, vyvolá se upozornění).
- Jaké modely lze využít:
  - statistické modely,
  - shlukování (k-means, k-medoids, k-nn clusters, ...),
  - klasifikační modely (bayesovské modely, SVM, neuronové sítě, ...)
  - pravděpodobnostní automaty, markovovy řetězce.
- Sbět trénovacích dat:
  - Netflow – informace o tocích (jaké IP adresy, jaké porty, jak dlouho tok trval, kolik dat se přeneslo, ...).
  - SNMP – informace o provozu (počet přenesených paketů, bytů, ...).
  - Úplné zachytávání paketů.
- Systém může fungovat pouze tehdy, pokud jsme schopni sestavit model běžného chování.

## Kapitola 23

# PDS – Sítě Peer-to-Peer: vlastnosti, chování, způsoby směrování. Strukturované a nestrukturované sítě.

### 23.1 Zdroje

- 08-p2p.pdf
- PDS\_2021-04-09.mp4

### 23.2 Architektura peer-to-peer sítí

- Peer-to-Peer (P2P) je alternativní architektura vůči client-server.
- Uzly (uživatelé) spolu komunikují napřímo. Každý uzel v síti má stejnou roli. Není zde žádný centrální bod, žádný uzel není nadřazený ostatním<sup>1</sup>.
- Jiný způsob adresování – adresování obsahem.
- Jiný způsob směrování – lokální rozhodování, specifická struktura sítí.
- Příklad: BitTorrent, Napster, Gnutella, Skype (dříve), Bitcoin, Bluetooth, instant messaging služby, ...

**Logická síť** Základem každé P2P sítě je tzv. logická síť (*overlay*), která je vystavěna na aplikační vrstvě TCP/IP. Tedy P2P sítě staví nad existující sítiovou infrastrukturou. Logická síť definuje způsob propojení uzlů, směrování, vyhledávání informací, ...

**Definice P2P sítě** Dynamický soubor nezávislých uzlů (peers), které jsou propojeny a jejichž zdroje (objekty) jsou k dispozici ostatním uzlům v této síti. Zdroje: výpočetní výkon, disková kapacita (soubory), zařízení (tiskárny). Sdílené zdroje jsou přímo přístupné všem uzlům, ty je nabízejí a zároveň využívají. Síť obsahuje prostředky pro připojení uzlu k síti, hledání a využití zdrojů, ....

### Typy P2P sítí

- Pravé (pure) – Odebrání libovolného uzlu ze sítě nemá vliv na ztrátu schopnosti sítě poskytovat služby (např. Bitcoin).

---

<sup>1</sup>Jsou výjimky.

- Hybridní – Pro svou činnost využívají centrální uzel pro poskytování části nabízených sítiových služeb. Centrální bod slouží k autentizaci, indexování, inicializaci uzlu, apod.

### Vlastnosti P2P sítí

- Samo-organizovatelnost:
  - Když se uzel připojí nebo odpojí, tak se síť přeskopí a funguje dále.
  - Decentralizovaná topologie, kde uzly spolupracují na jejím vytvoření a udržování.
  - Každý uzel je zodpovědný za svůj lokální stav a část informací (zdrojů).
  - Uzly mají částečný pohled na topologii sítě (znají své nejbližší sousedy).
- Autonomní chování (samoriditelnost):
  - Uzly se chovají dle svého nejlepšího rozhodování (uzel pouze konzumuje zdroje, ale nechce poskytovat).
  - Rozhodování je lokální a nepredikovatelné ⇒ má vliv na topologii sítě, směrování, rozmístění objektů.
  - Uzly se mohou chovat zlomyslně.
  - Problém s ověřováním identity uzlů a důvěryhodností (decentralizované řízení).
- Spolehlivost:
  - Spolehlivost sítě roste s redundancí uzlů a informací.
  - Redundance objektů, kopie jsou umístěny ve více uzlech.
- Životnost uzlu:
  - Doba životnosti uzlu je neodhadnutelná ⇒ problém s garancí služby.
  - Závisí na subjektivním lokální rozhodnutí.

|                           | Klient – server      | Peer-to-Peer                         | <i>Výhody/nevýhody P2P</i>                         |
|---------------------------|----------------------|--------------------------------------|--|
| <b>Směr provozu</b>       | Asymetrický          | Symetrický                           | <i>vs. xDSL, kabelový modem</i>                    |
| <b>Topologie sítě</b>     | Stabilní             | Dynamická                            | <i>Problém spolehlivosti</i>                       |
| <b>Robustnost</b>         | Centrální bod        | Distribuce zdrojů                    | <i>Kritický počet účastníků</i>                    |
| <b>Rozšiřitelnost</b>     | Náročné              | Součást návrhu                       | <i>Neomezený růst sítě</i>                         |
| <b>Bezpečnost</b>         | Velký důraz          | Problematické                        | <i>Chybí odpovědná autorita</i>                    |
| <b>Správa a řízení</b>    | Centralizovaný model | Každý uživatel spravuje vlastní uzel | <i>Samo-organizovaná síť</i>                       |
| <b>Poskytované zdroje</b> | Omezené možnosti     | Dynamicky rostoucí počet zdrojů      | <i>Sdílení výpočetního prostoru, paměti, apod.</i> |
| <b>Kvalita služeb</b>     | Garantovaná          | Nelze zajistit                       | <i>Dynamicky se mění</i>                           |

Obrázek 23.1: Srovnání vlastností P2P a klient-server architektur.

## Referenční model P2P

- Mějme množinu uzlů  $P$ , množinu zdrojů  $R$  a množinu identifikátorů  $I$ .
- Struktura logické sítě:
  - mapování zdrojů:  $F_R : R \rightarrow I$ ,
  - mapování uzlů:  $F_P : P \rightarrow I$ .
  - Množina uzlů  $P$  zpřístupňuje zdroje  $R$  v rámci jmenného prostoru  $I$  pomocí  $F_R$  a  $F_P$ .
- Decentralizovaná správa jmenného prostoru:
  - $M : I \rightarrow 2^P$
  - Příklad: uzel potřebuje zjistit, kdo má konkrétní zdroj.
- Metrika blízkosti:
  - $d : I \times I \rightarrow \mathbb{R}$
  - Příklad: uzel chce konkrétní zdroj, vyhledá všechny uzly, které ho poskytují a na základě metriky blízkosti vybere nejbližšího.

## Geometrie sítě a směrování

- Geometrie (topologie).
  - Dynamická, uzly se připojují a odpojují.
  - Množina uzlů  $P$  zpřístupňuje zdroje  $R$  v rámci jmenného prostoru  $I$  pomocí  $F_R$  a  $F_P$ .
- Směrování.
  - Každý uzel zná své sousedy (uzly a hrany k nim) pomocí relace sousedství:  $N : P \rightarrow 2^P$  (uzel  $\rightarrow$  jeho sousedi).
  - Mějme předávání zprávy  $route(p, m, i)$ , kde hledáme cestu pro zprávu  $m$ , k uzlu  $p$ , který spravuje zdroj  $i$ . Směrování: kterému sousedovi mám zprávu předat?
  - Distribuovaný proces nalezení cesty v síti P2P na základě lokálních znalostí.
  - Směrovací funkce:  $R : P \times I \rightarrow 2^P$ .
  - Směrovací tabulka: každý uzel má svoji, kterou si postupně plní a aktualizuje.

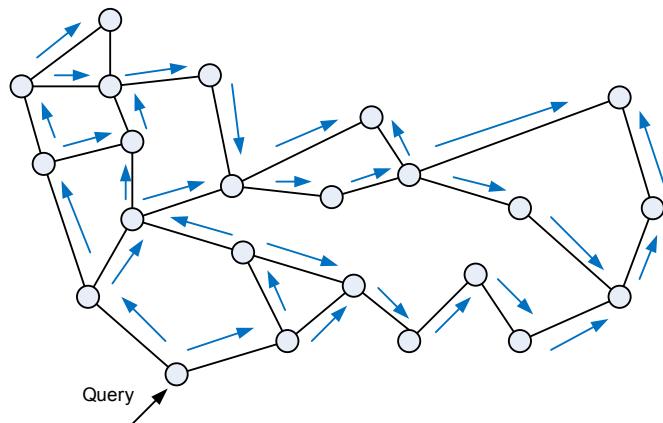
## 23.3 Nestrukturované sítě

- Neexistuje struktura uložení informace (zdroj (objekt) je umístěn na náhodném uzlu).
- Uzel si vyměňuje zprávy se svými sousedy (dotaz na vyhledávání konkrétního objektu).
- Když uzel hledá objekt, neví, jestli se přibližuje, dostává odpovědi pouze ano-ne.
- Jak v takovém systému směrovat (jak najít objekt skrze identifikátor)?

### 23.3.1 Záplava (*flooding*)

- Uzel pošle dotaz na objekt všem svým sousedům.

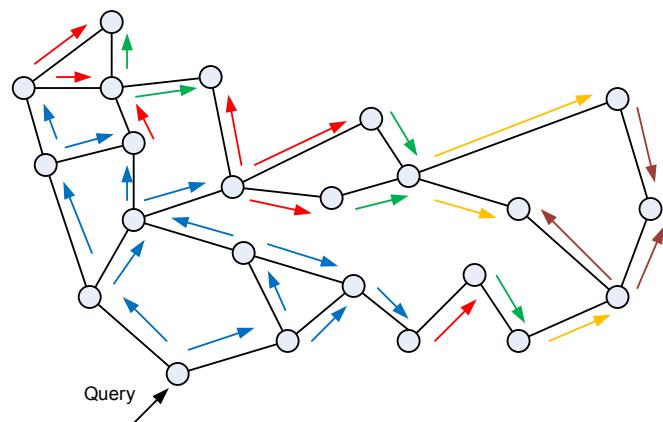
- Pokud soused má objekt, pošle zpět odpověď.
- Pokud nemá, pošle zprávu svým sousedům.
- Lze omezit pomocí TTL ve zprávě (zamezuje zacyklení a zahlcení sítě).



Obrázek 23.2: Záplava.

### 23.3.2 Rozšiřující se kruh (*expanding ring*)

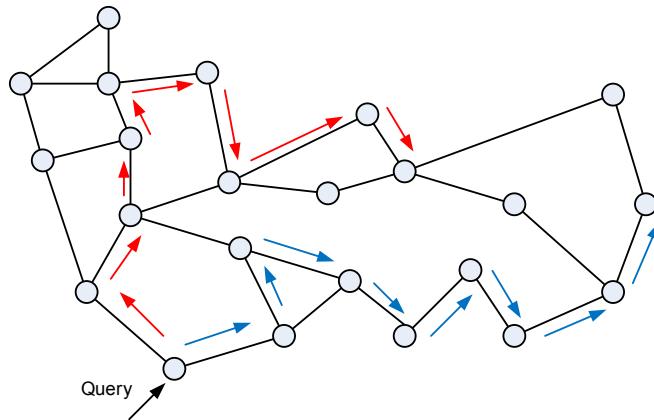
- Uzel pošle dotaz na objekt všem svým sousedům s malým TTL.
- Pokud objekt je nalezen, hledání končí.
- Pokud objekt není nalezen, zvýší se TTL a dotaz se pošle znovu.
- Redukuje počet zpráv v síti.



Obrázek 23.3: Rozšiřující se kruh.

### 23.3.3 Náhodný průchod (*random walk*)

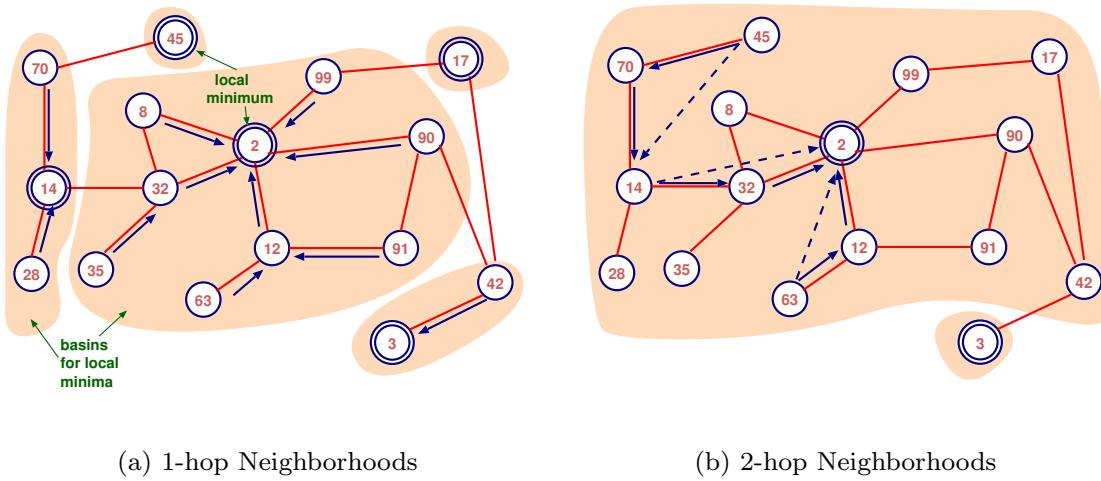
- Uzel pošle dotaz na objekt náhodně vybraným sousedům (i více zároveň).



Obrázek 23.4: Náhodný průchod.

### 23.3.4 Hledání lokálního minima (*local minimum search*)

- Není pro čistě nestrukturované sítě (přidává lokální vzdálenost).
- Máme množinu uzlů identifikovaných hodnotou  $x$  a množinu objektů s identifikátorem  $w$ . Úkolem je umístit objekty do sítě uzlů tak, abychom je mohli najít rychle a spolehlivě, tj. jméno uzlu  $x$  by mělo být co nejblíže jménu ukládaného objektu  $w$ .
- Při hledání používáme metriku vzdálenosti uzlu  $x$  od objektu  $w$  –  $d(x, w)$ .
- Uzel  $u$  je lokálním minimem pro objekt, pokud je jeho ID nejbližší k ID objektu mezi jeho sousedy do vzdálenosti  $h$  kroků.



Obrázek 23.5: Příklad hledání lokálního minima. Uzly jsou značeny vzdáleností od objektu.  $2 - hop$  – uzel zná sousedy sousedů a zapojuje je do hledání.

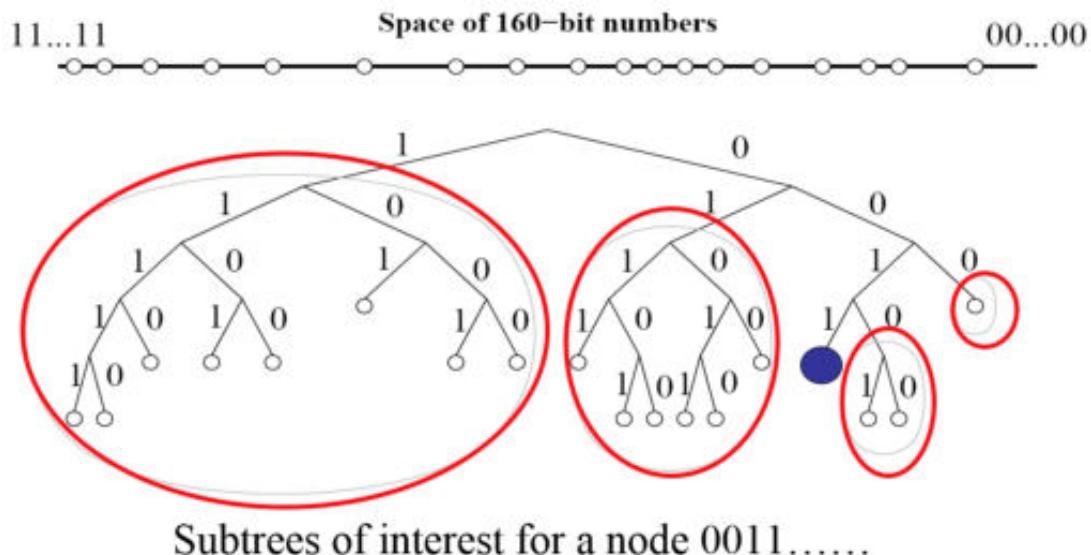
## 23.4 Strukturované sítě

- Kombinují geometrické struktury a směrování.
- Distribuované směrovací algoritmy (metriky: shoda prefixu, euklidovská či lineární vzdálenost, XOR, ...).

- Struktura sítě odpovídá uložení zdrojů.
- Jak směrovat?

### 23.4.1 Kademlia – Distribuovaná hašovací tabulka (DHT)

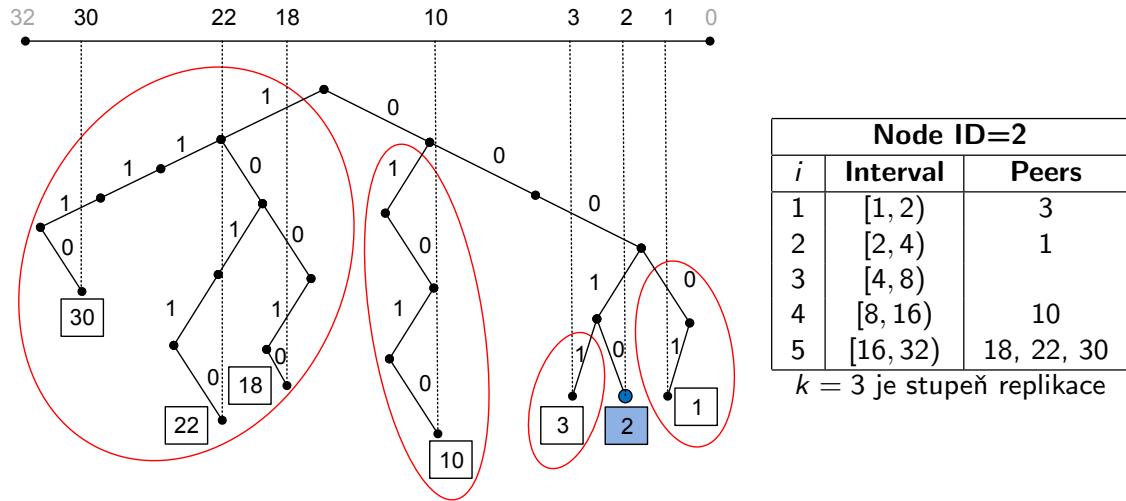
- Identifikátory uzlů i souborů jsou hash.
- Metrika blízkosti: bitový XOR ( $d(a, b) = a \oplus b$ ).
- Pro směrování používá distribuovanou hašovací tabulku (DHT), která obsahuje informace o dalších uzlech a souborech.
  - Myšlenka: znám dobře svoje okolí, čím dálé jsem, tím méně znám. Ale mám tam kontakty, kterých se můžu ptát a které mají lepší lokální informace.
  - Každý uzel má uloženou DHT ze své perspektivy.



Obrázek 23.6: DHT se 160 bitovým jmenným prostorem. Uzly jsou umístěny ve stromu podle prefixů.

- **Směrovací tabulka** vzniká dynamicky, tím že přijde zpráva od jiného uzlu (spočítá se vzdálenost a uzel se přidá do tabulky).
  - Pokud daný řádek obsahuje méně než k položek (počet replikací), tak se uzel přidá.
  - Pokud je plný, tak zkontroluji, jestli jsou uzly dostupné, případně aktualizuju.
- Tento princip využívá P2P sítě **BitTorrent** (bud' využívá DHT a nebo koncept *trackers*).
  - *tracker* – Uzel, který má informace o tom, který uzel obsahuje jaké zdroje.
  - *torrent* – Soubor s informací o sdílených souborech a s odkazem na tracker.
  - BitTorrent je pak aplikační protokol pro distribuci souborů nad TCP.

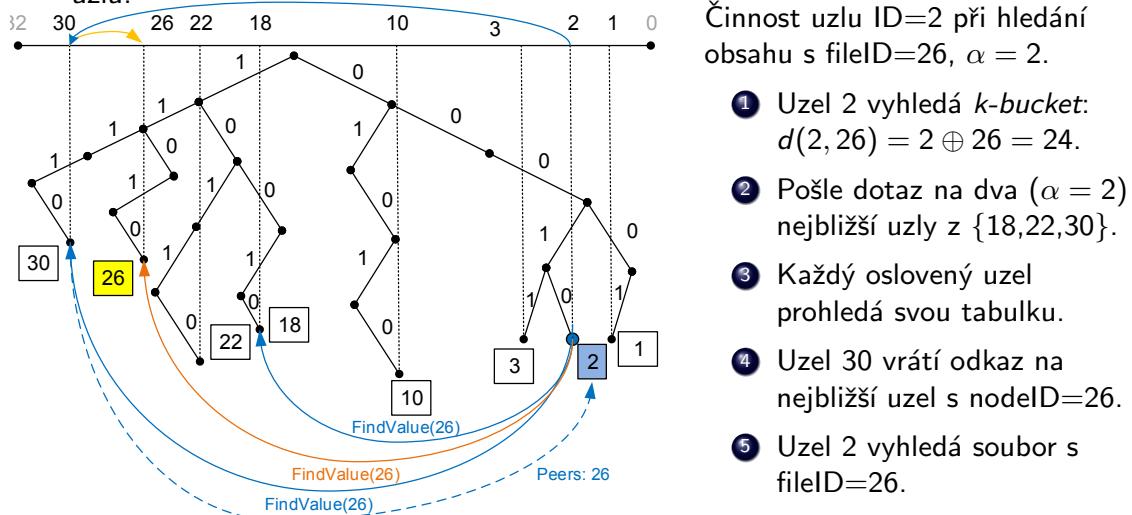
- Jmenný prostor o velikosti  $N = 32$ ,  $k=3$ , délka ID je 5 bitů.
- Uzly ve stejném podstromu mají stejný prefix i stejnou vzdálenost od daného uzlu.
- Např.  $d(2, 3) = 2 \oplus 3 = 1$ ,  $d(10 \oplus 2) = 8$ ,  $d(22 \oplus 2) = 20$ ,  $d(30 \oplus 2) = 28$ , apod.



- Řádek  $i$  obsahuje uzly podstromu v dané vzdálenosti  $[2^i, 2^{i+1})$  od daného uzlu.

Obrázek 23.7: Konkrétní příklad DHT pro 5 bitový jmenný prostor. Napravo je směrovací tabulka uzlu 2. *Interval* je interval vzdáleností a *Peers* uzly co se v této vzdálenosti vyskytují.

- PING(nodeID): kontroluje, zda daný uzel je připojen.
- STORE(fileID, nodeID): uloží do daného uzlu fileID a nodeID, který ho nabízí.
- FIND\_NODE(nodeID): daný uzel vrátí nejbližší uzly k uzlu nodeID.
- FIND\_VALUE(fileID): vrátí adresu uzlu obsahující soubor nebo seznam nejbližších uzlů.



Obrázek 23.8: Příklad vyhledávání v DHT. Uzel 2 hledá zdroj 26. XOR vzdálenost 2 a 26 je 24. Proto se ptá uzlů, které má ve vzdálenosti  $\langle 16, 32 \rangle$ . Jeden uzel, podá přesnější informaci (uzel s ID 26). Uzel 2 se následně zeptá uzlu 26.