14

Porovnávání jazyků a redukce automatů používaných při filtraci síťového provozu

Vojtěch Havlena

xhavle03@stud.fit.vutbr.cz

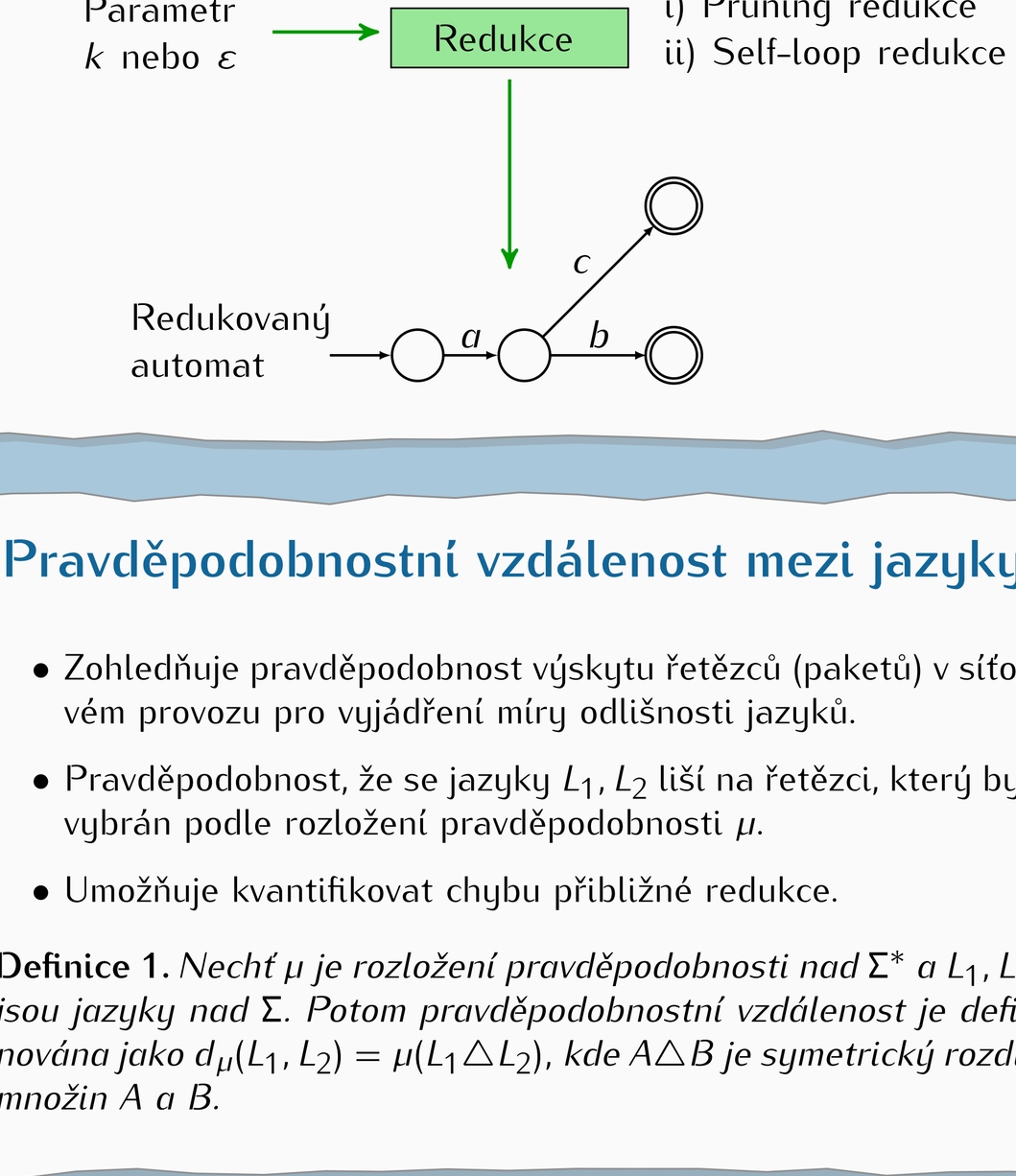
Excel@FIT 2017

Motivace

- HW akcelerované filtrování síťového provozu na základě konečných automatů.
- **Problém:** velikost automatu uloženého v HW.
- Klasické redukční techniky zachovávající jazyk (DFA minimalizace, simulační redukce) nemusí být dostatečné.

- **Přibližné redukce**
  - Vyšší stupneň redukce
  - Nezachovávají jazyk
  - Garance maximální chyby (vůči vstupnímu provozu)

Architektura



Pravděpodobnostní vzdálenost mezi jazyky

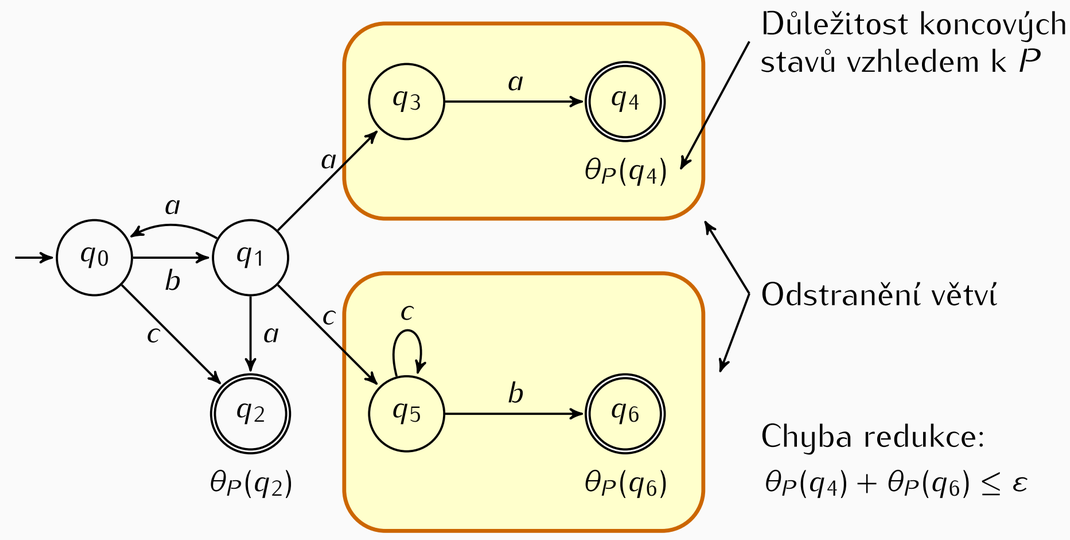
- Zohledňuje pravděpodobnost výskytu řetězců (paketů) v síťovém provozu pro vyjádření míry odlišnosti jazyků.
- Pravděpodobnost, že se jazyky L_1, L_2 liší na řetězci, který byl vybrán podle rozložení pravděpodobnosti μ.
- Umožňuje kvantifikovat chybu přibližné redukce.

Definice 1. *Nechť μ je rozložení pravděpodobnosti nad Σ^* a L_1, L_2 jsou jazyky nad Σ. Potom pravděpodobnostní vzdálenost je definována jako $d_\mu(L_1, L_2) = \mu(L_1 \triangle L_2)$, kde $A \triangle B$ je symetrický rozdíl množin A a B.*

Metody redukce

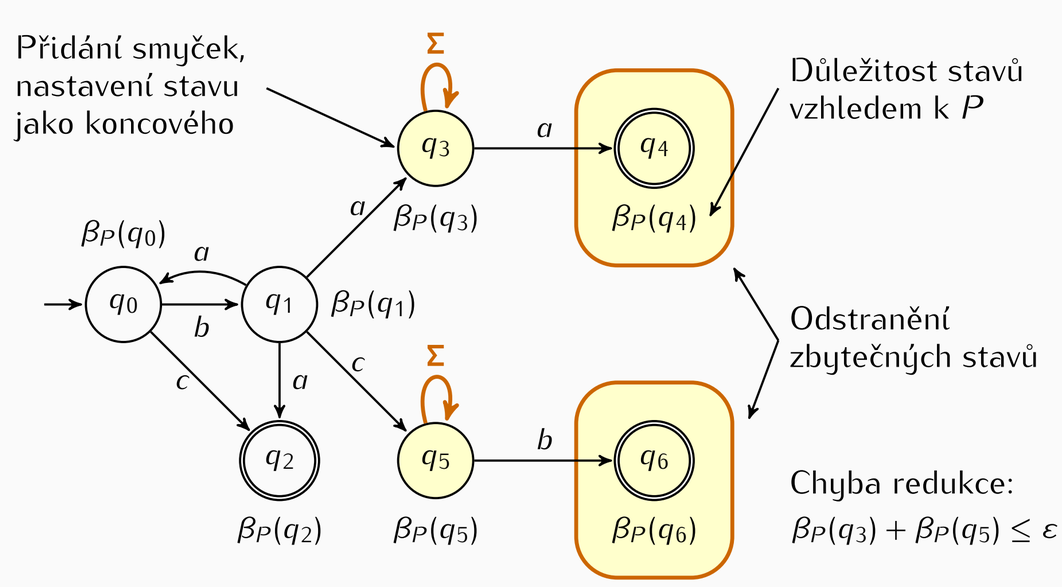
Pruning redukce

- Odstranění méně důležitých větví automatu na základě pravděpodobnostního automatu P.



Self-loop redukce

- Výběr stavů pro přidání smyček na základě pravděpodobnostního automatu P, následné odstranění zbytečných stavů.



Experimenty

1. Učení pravděpodobnostního automatu z vzorku provozu.
2. Redukce automatů popisujících útoky/protokoly na základě naučeného automatu.
3. Vyhodnocení chyby na reálném vzorku provozu.

k	Počet stavů	Chyba na vzorku provozu	Teoretická chyba
0.0	1	1.0	1.0
0.2	4	0.00861	0.03041
0.5	9	0.0	4.04245e-10
0.7	12	0.0	1.93657e-12
1.0	16	0.0	0.0

Tabulka. Self-loop redukce automatu `info.rules`.

- Prvotní experimenty na několika automatech:
  - **Redukce: 70%**
  - **Chyba: <3%**