

Úvod do SHO



Výkonnost a spolehlivost – KIV/VSP

Richard Lipka

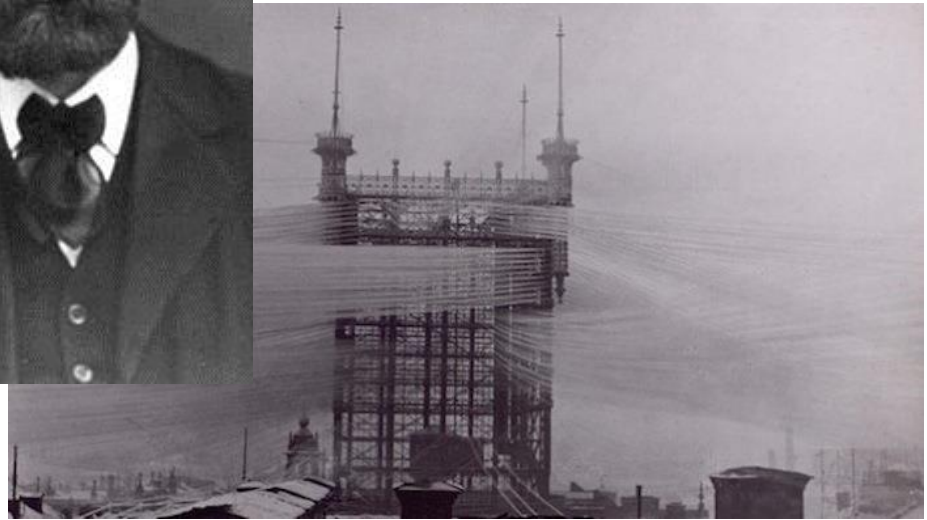
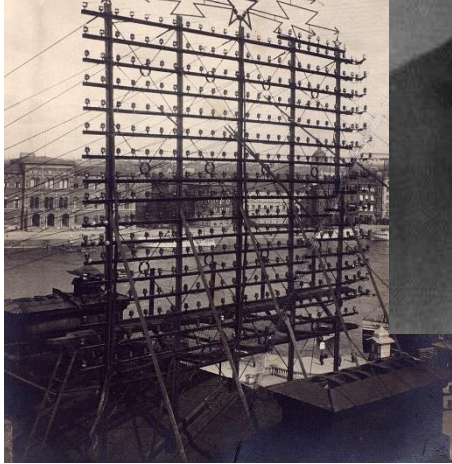
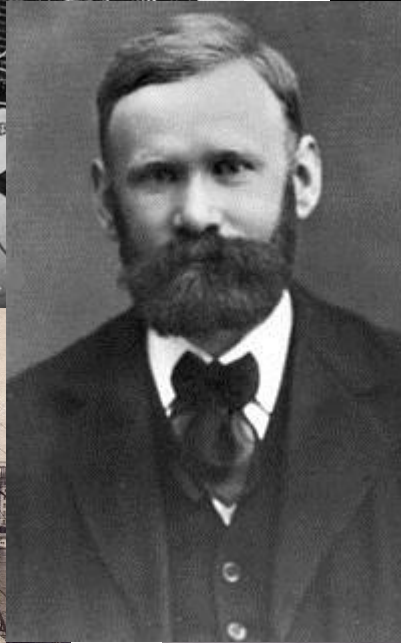
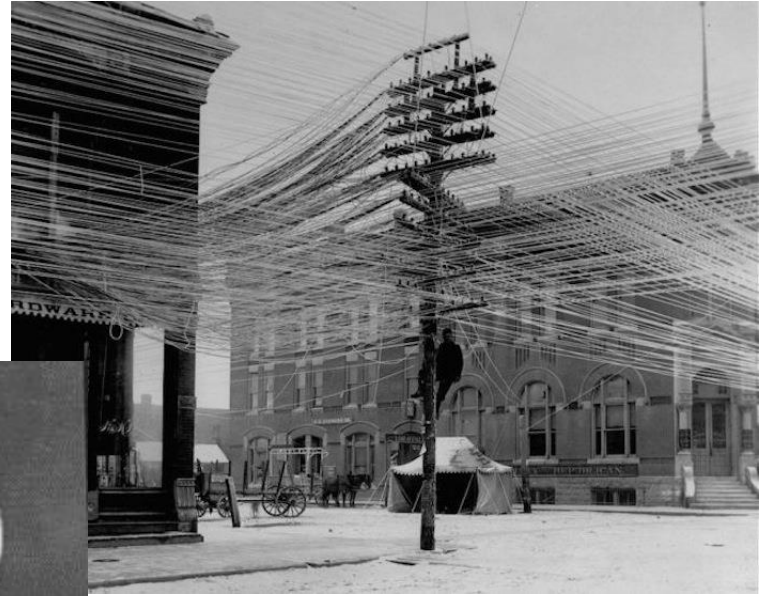
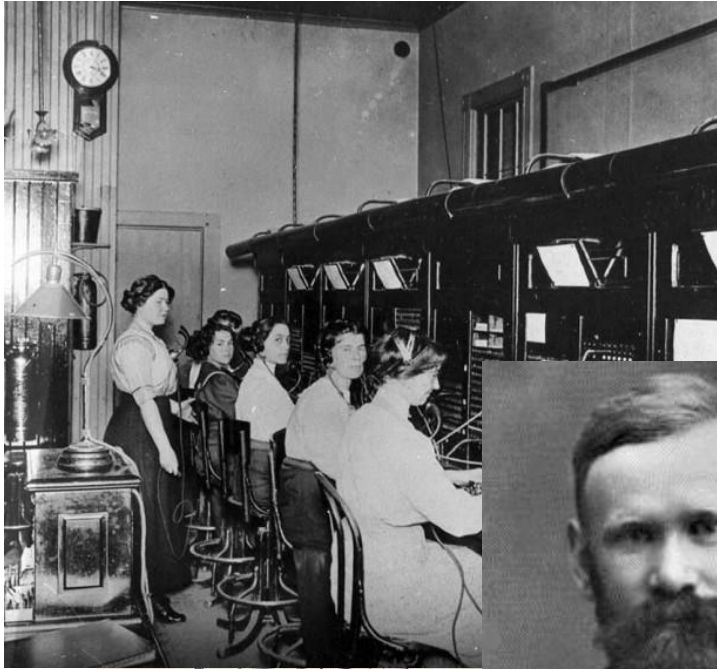
13.10.2015



Systemy hromadné obsluhy (*Queueing theory*)

- Modelování systémů, které obsluhují větší množství požadavků
 - Telekomunikační systémy
 - Řízení dopravy
 - Plánování procesů v OS
 - Návrh výrobních linek
 - Krizový management
 - Návrh míst, kde se čeká (obchody, úřady, nemocnice ...)
- Predikce výkonu takových systémů
- První modely ze začátku 20. století (telefonní ústředna v Kodani – *Erlang, 1909*)
 - podobně staré jako **Markovské modely**
 - Řada vztahů odvozena právě přes markovské modely





VSP - Úvod do SHO I



SHO - základy

- Základní koncept:
 - *Systém poskytuje službu*
 - Realizována *kanálem obsluhy / serverem*
 - *Klienti posílají požadavky / zahajují transakce*
 - *Požadavky jsou buď obsluhovány, nebo čekají ve frontách*
- Teorie založena na statistice → funguje jen pro dostatečný počet požadavků
 - Abstrahuje od konkrétní realizace služby
 - Sleduje jen časové posloupnosti příchodů požadavků a jejich obsluh
→ lze určit základní výkonové charakteristiky
- Zjednodušení umožňuje konstruovat SHO model v uzavřeném tvaru (jako vzoreček)
(složité SHO je třeba řešit simulačně, ne vždy je možné nebo praktické najít uzavřený tvar)

VSP - Úvod do SHO I



Příklady



| System | Klient | Server | Požadavek |
|--------------|-----------------|------------------------------|--------------------------------|
| Dopravní síť | Vozidlo | Křižovatka | Průjezd křižovatkou |
| Google | Hledající | Vyhledávací stroj a DB | Vyhledání stránky |
| Letiště | Letadlo | Runway | Odlet nebo přilet |
| Úřad | Občan | Přepážka | Vydání dokladu |
| Menza | Hladový student | Výdejní okénko / pult / kasa | Vydání jídla / zaplacení jídla |
| Myčka aut | Vozidlo | Mycí linka | Umytí vozidla |

VSP - Úvod do SHOI

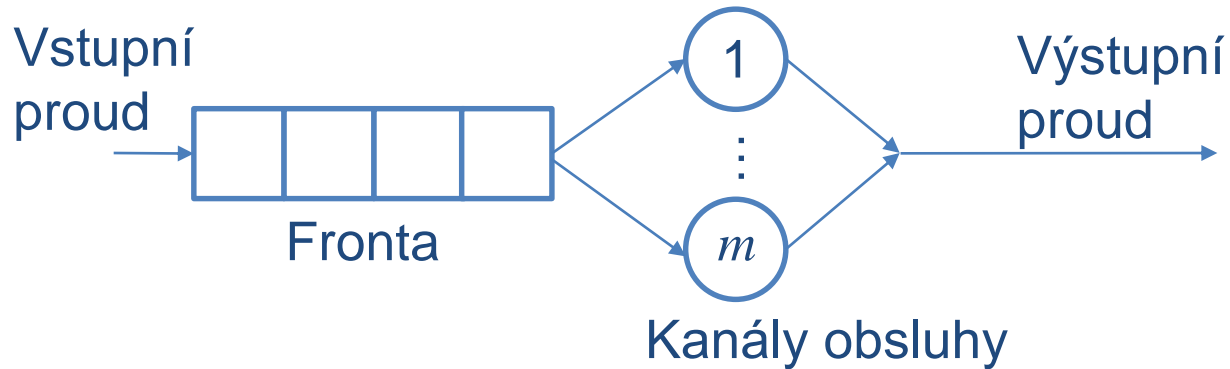


Základní problémy

- Příliš mnoho požadavků na systém
 - Požadavky čekají ve frontě příliš dlouho (fronta může narůstat do nekonečna)
 - Požadavky nemají kde čekat a jsou zahazovány
 - potřebuji zvýšit kapacitu systému (přidat zdroje) – relativně snadné v cloudových řešeních
- Příliš málo požadavků na systém
 - Systém není vytížen – většinu času nic nedělá (ale stále spotřebovává zdroje)
 - mohu zdroje využít nějak jinak (nebo zrušit)



Elementární SHO



- Stav systému: počet požadavků v něm
- Potřebujeme znát:
 - Charakteristiku vstupního proudu
 - Chování fronty
 - Charakteristiku obsluhy požadavků
- Předpokládáme stacionární režim činnosti
 - Charakteristiky se nemění v čase (podobně jako u Markovských modelů)
 → modelujeme *ustálený provoz*, ne *přechodový děj*

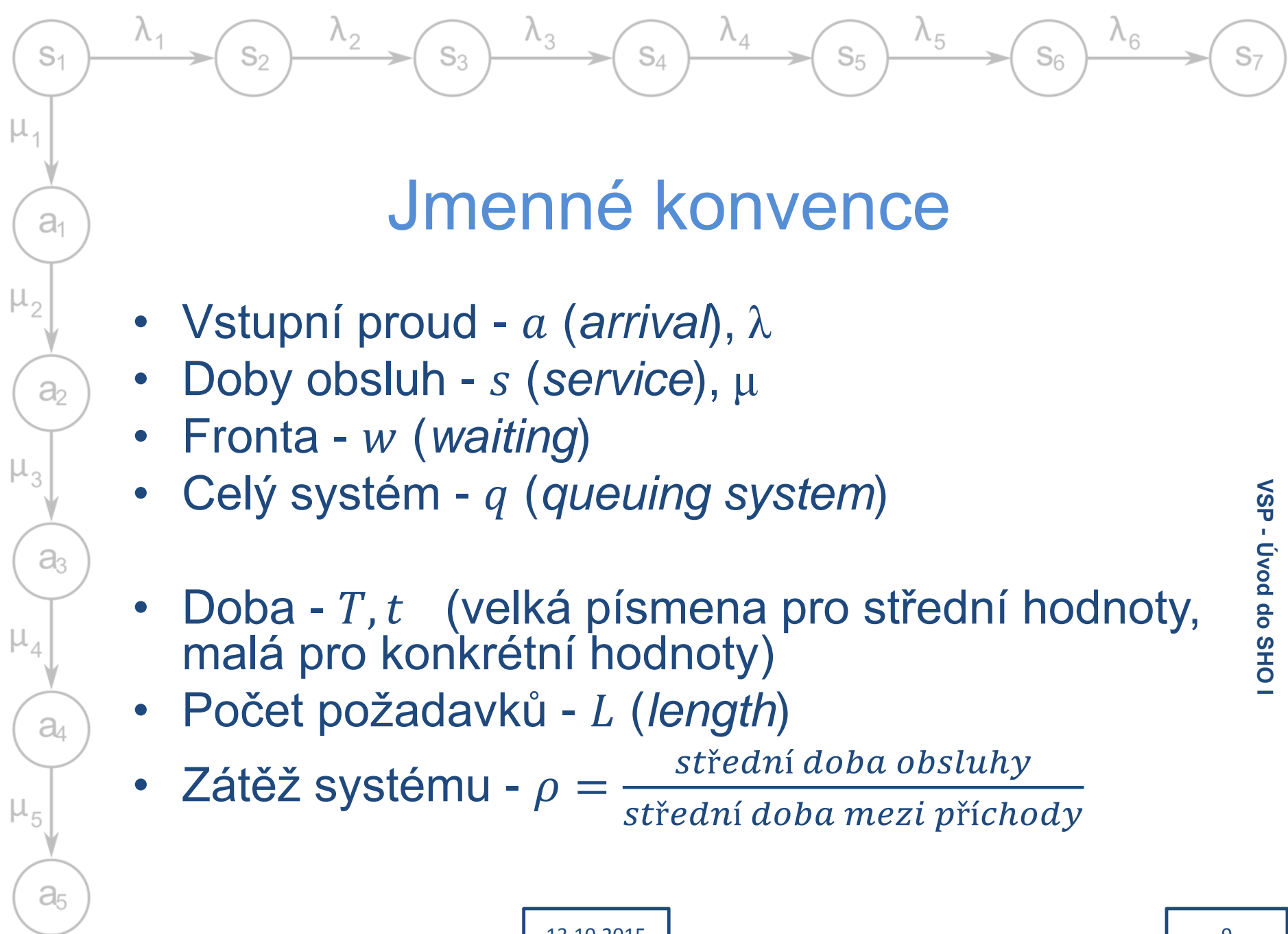
VSP - Úvod do SHO I



Zdroje požadavků



- Vkládají požadavky do fronty
 - Abstraktní, obvykle bez ekvivalentu v reálném světě, hodí se hlavně pro simulace
- Lze dělit podle počtu požadavků
 - *Omezené zdroje*
 - Předem omezená množina požadavků
→ vstupní proud závisí na stavu SHO
 - Např. modelování procesů v kritické sekci
 - *Neomezené zdroje*
 - „Nekonečné“ množství požadavků
→ vstupní proud nezávisí na stavu SHO
 - Např. požadavky na webový server



Jmenné konvence

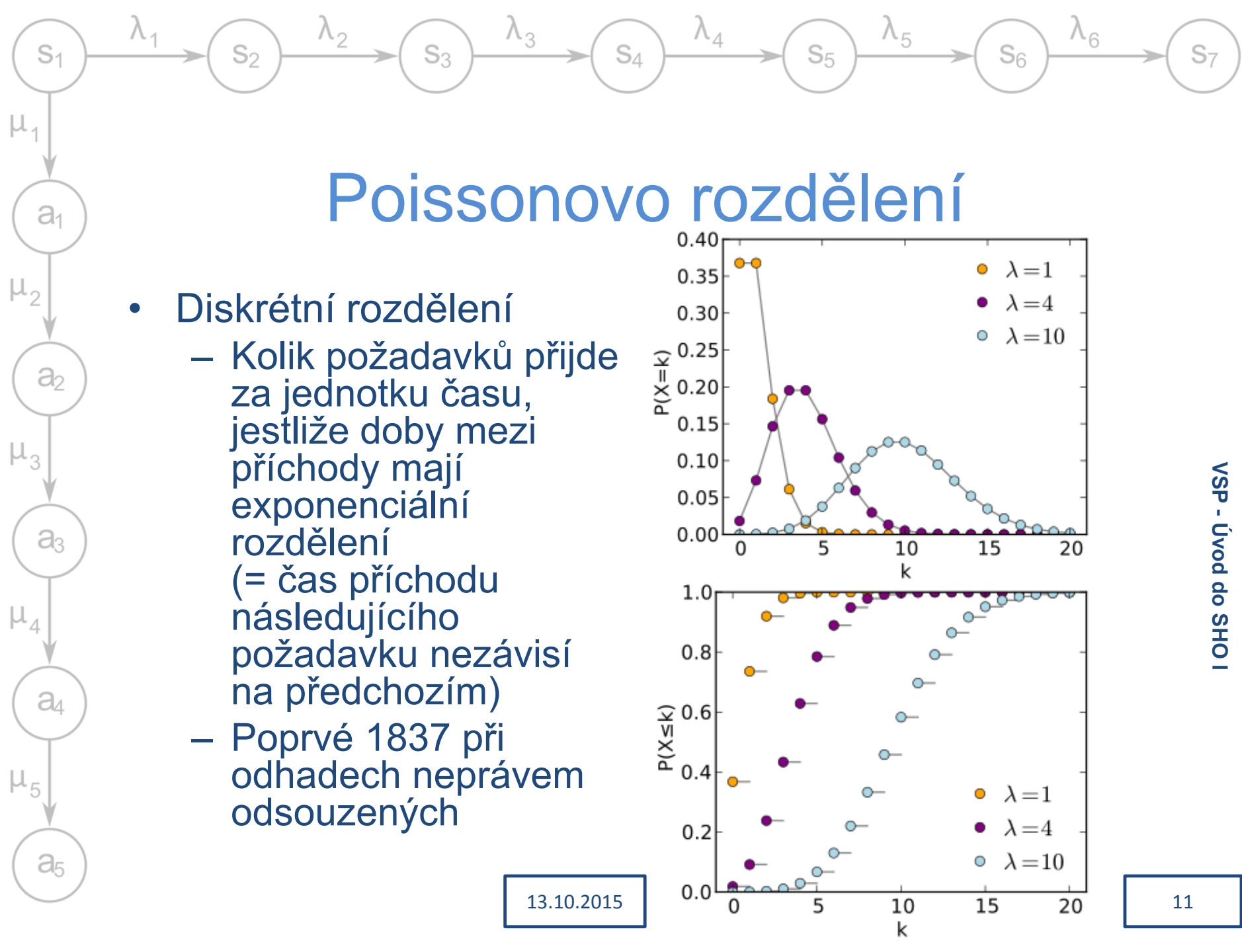
- Vstupní proud - a (*arrival*), λ
- Doby obsluh - s (*service*), μ
- Fronta - w (*waiting*)
- Celý systém - q (*queueing system*)
- Doba - T, t (velká písmena pro střední hodnoty, malá pro konkrétní hodnoty)
- Počet požadavků - L (*length*)
- Zátěž systému - $\rho = \frac{\textit{střední doba obsluhy}}{\textit{střední doba mezi příchody}}$



Vstupní proud

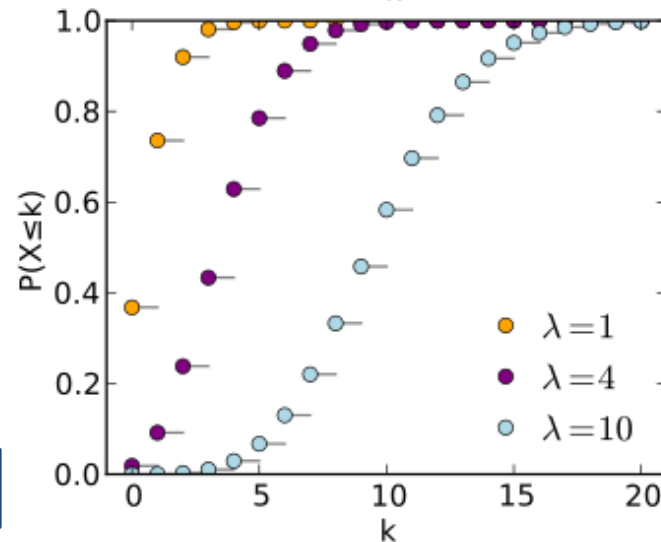
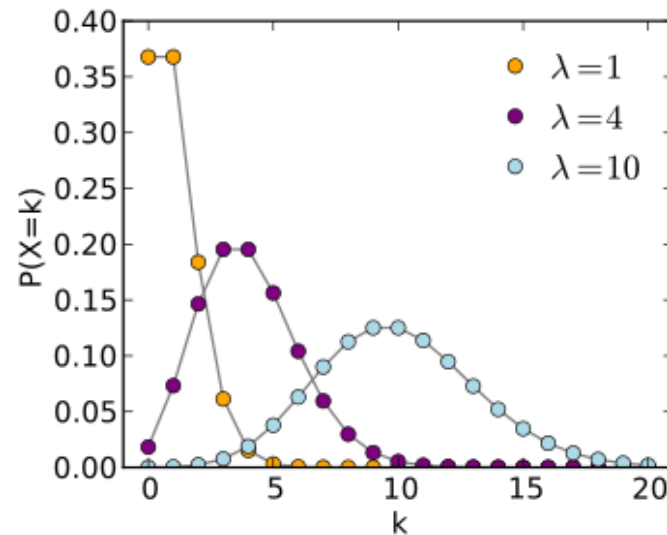
- Časová posloupnost s jakou požadavky do systému vstupují - $\{t_1 < t_2 < t_3 \dots\}$
 - Popsán jako $\tau_k = t_k - t_{k-1}$ pro $k \geq 1$ – interval mezi příchody (*inter-arrival time*)
 - Pokud jsou jednotlivé hodnoty τ_k statisticky nezávislé a mají stejné rozdělení, považovány za realizaci veličiny τ
- Obvyklý popis: $F_a(t) = P\{\tau \leq t\}$ nebo odpovídající $f_a(t) = (F_a(t))'$
- Typické vstupní proudy:
 - *Poissonovský* – exponenciální rozdělení pro doby mezi příchody požadavků (v každém okamžiku stejná pravděpodobnost příchodu dalšího požadavku – nezáleží na tom jak dlouhá doba uběhla od posledního)
 - Gaussovský
 - Rovnoměrný – bez náhody, stejné intervaly

VSP - Úvod do SHO I

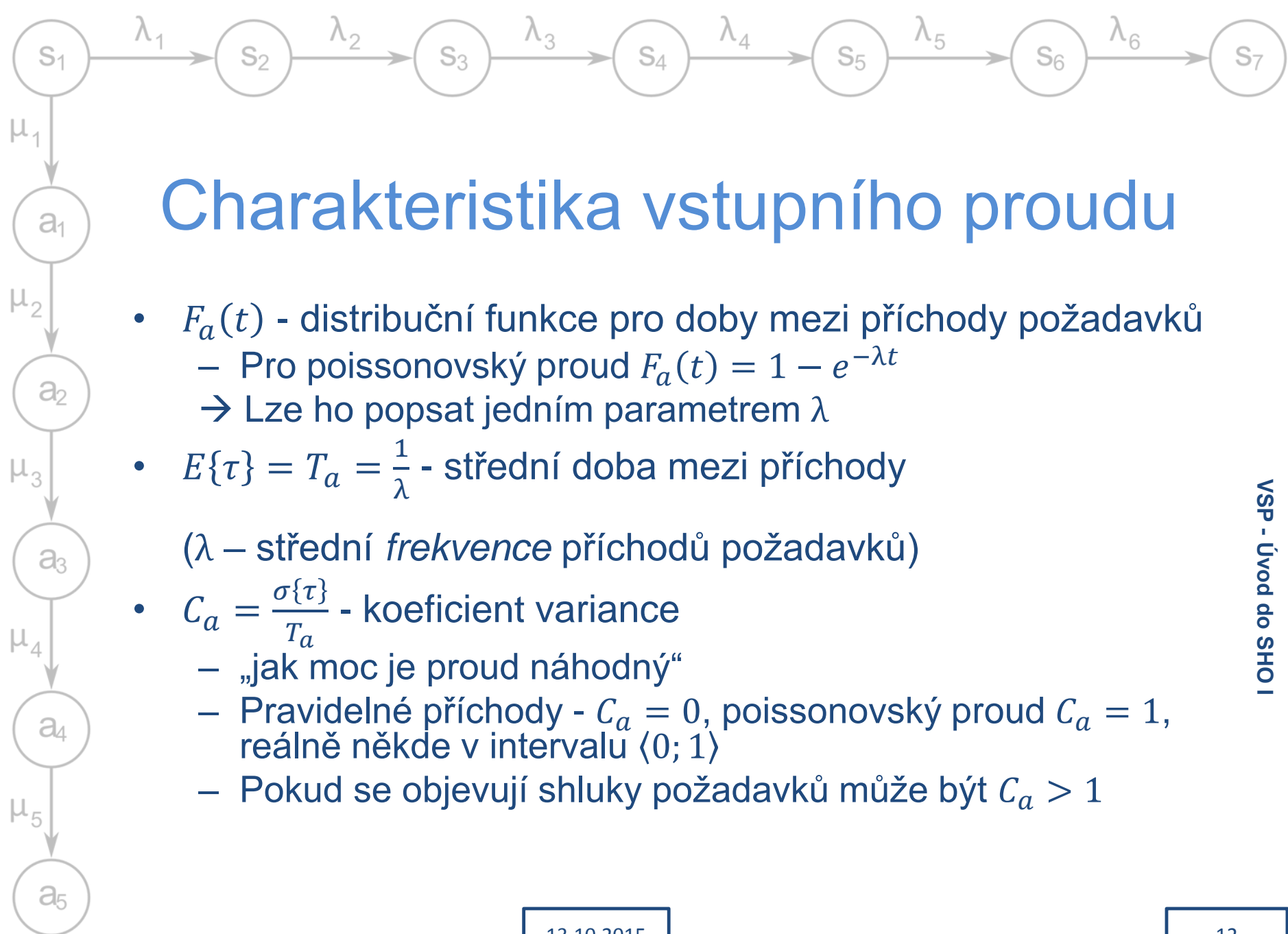


Poissonovo rozdělení

- Diskrétní rozdělení
 - Kolik požadavků přijde za jednotku času, jestliže doby mezi příchody mají exponenciální rozdělení (= čas příchodu následujícího požadavku nezávisí na předchozím)
 - Poprvé 1837 při odhadech neprávem odsouzených

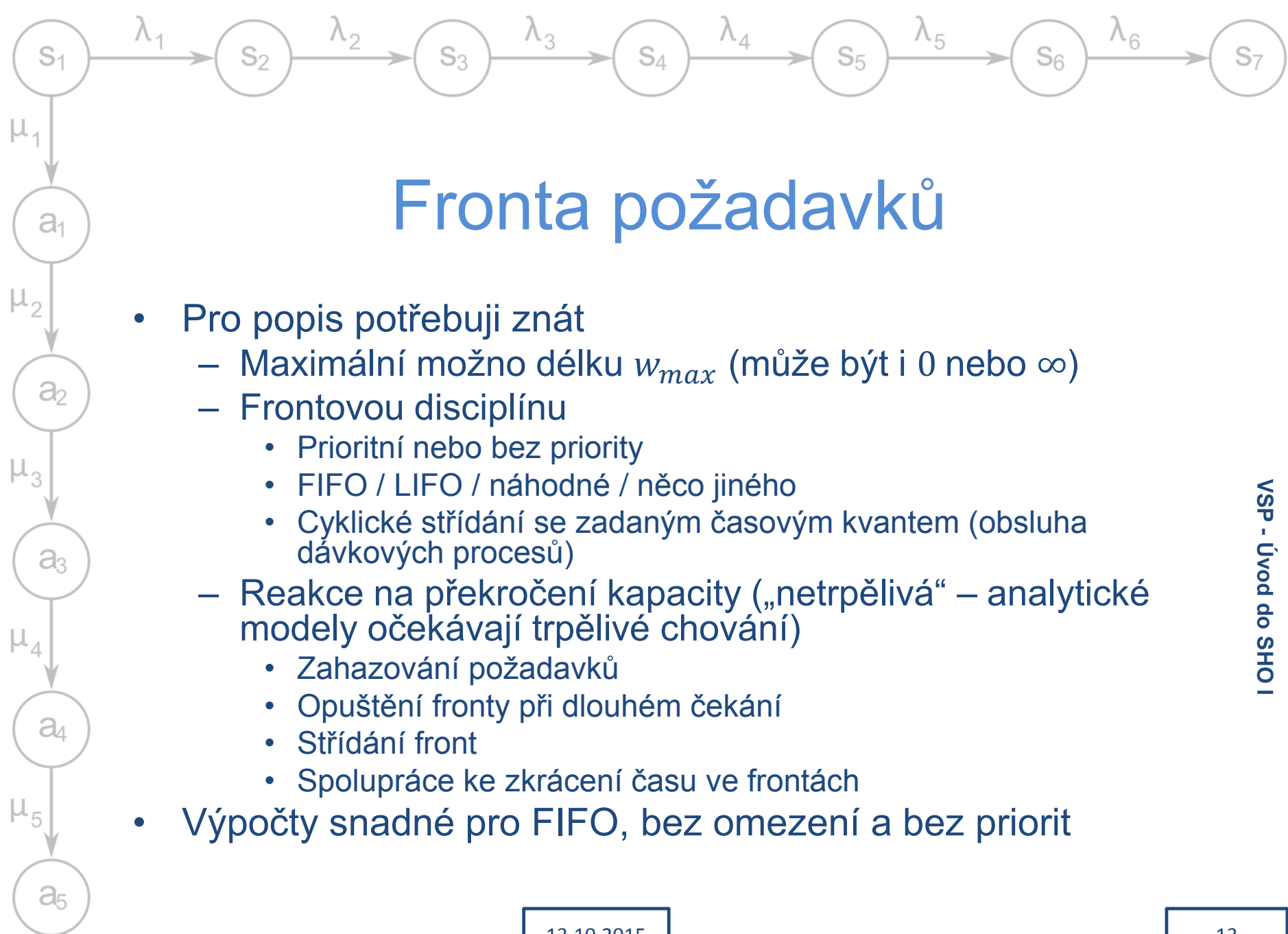


VSP - Úvod do SHOI



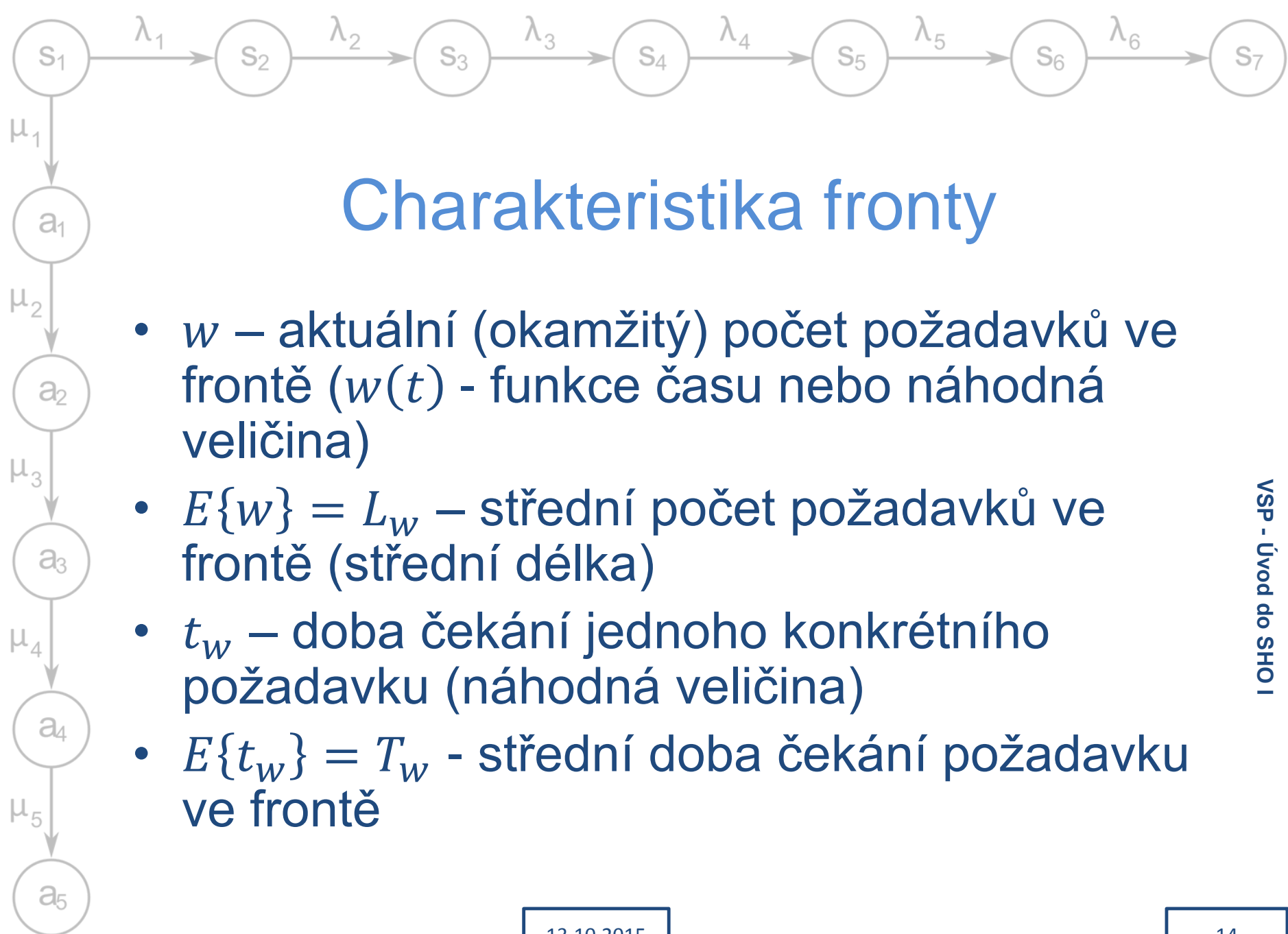
Charakteristika vstupního proudu

- $F_a(t)$ - distribuční funkce pro doby mezi příchody požadavků
 - Pro poissonovský proud $F_a(t) = 1 - e^{-\lambda t}$
 - Lze ho popsat jedním parametrem λ
- $E\{\tau\} = T_a = \frac{1}{\lambda}$ - střední doba mezi příchody
 (λ – střední *frekvence* příchodů požadavků)
- $C_a = \frac{\sigma\{\tau\}}{T_a}$ - koeficient variance
 - „jak moc je proud náhodný“
 - Pravidelné příchody - $C_a = 0$, poissonovský proud $C_a = 1$, reálně někde v intervalu $\langle 0; 1 \rangle$
 - Pokud se objevují shluky požadavků může být $C_a > 1$



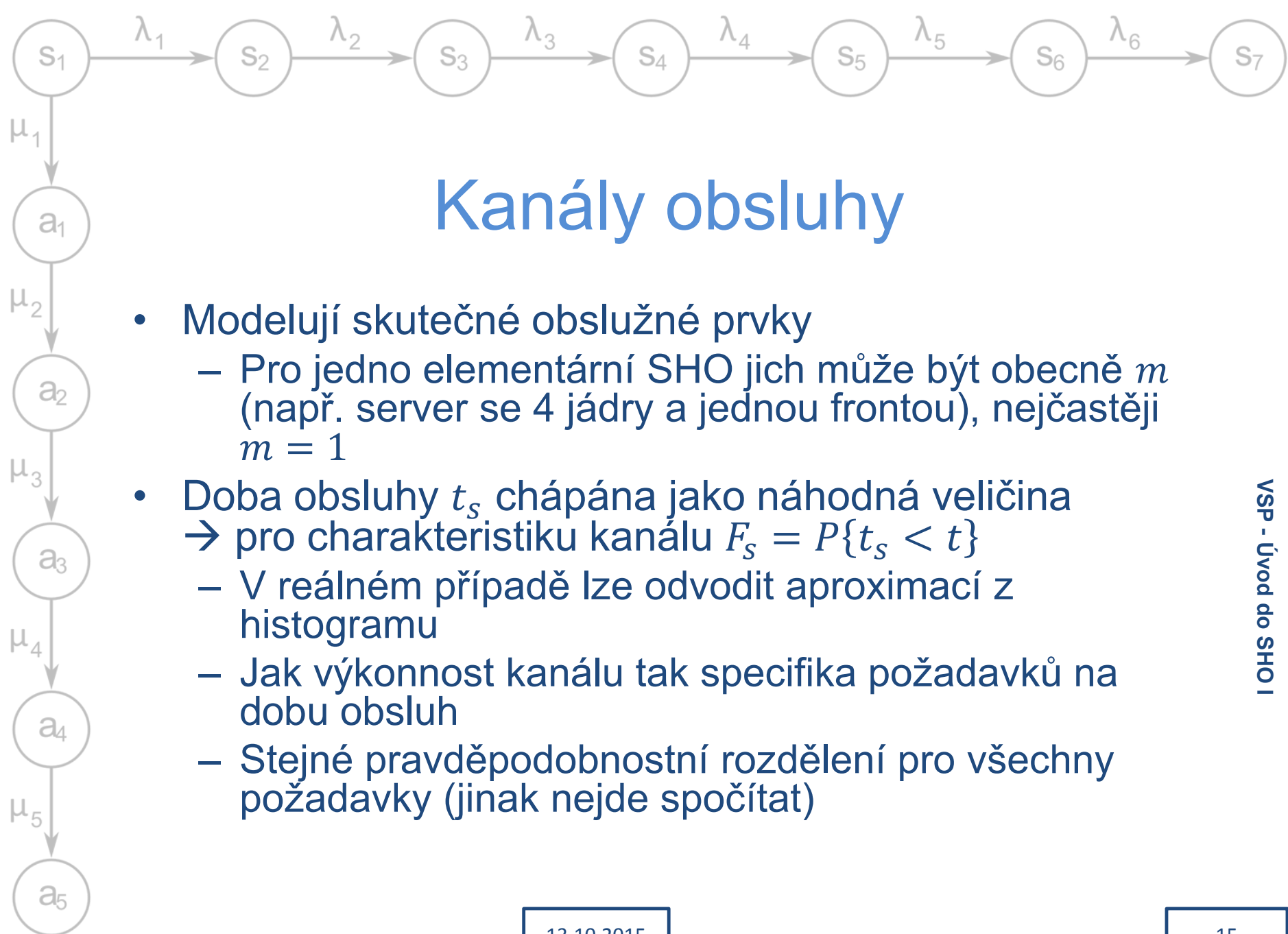
Fronta požadavků

- Pro popis potřebuji znát
 - Maximální možno délku w_{max} (může být i 0 nebo ∞)
 - Frontovou disciplínu
 - Prioritní nebo bez priority
 - FIFO / LIFO / náhodné / něco jiného
 - Cyklické střídání se zadaným časovým kvantem (obsluha dávkových procesů)
 - Reakce na překročení kapacity („netrpělivá“ – analytické modely očekávají trpělivé chování)
 - Zahazování požadavků
 - Opuštění fronty při dlouhém čekání
 - Střídání front
 - Spolupráce ke zkrácení času ve frontách
- Výpočty snadné pro FIFO, bez omezení a bez priorit



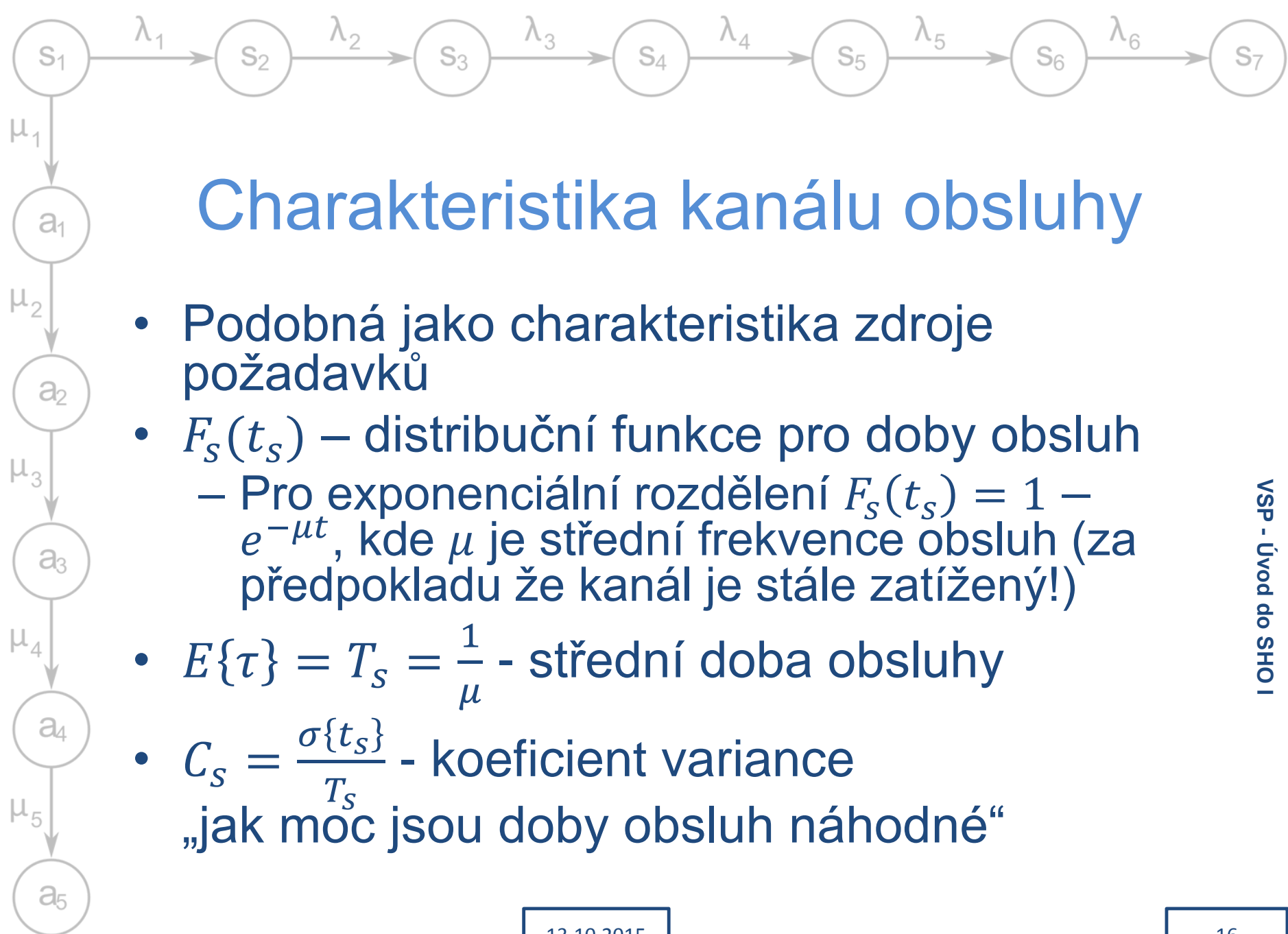
Charakteristika fronty

- w – aktuální (okamžitý) počet požadavků ve frontě ($w(t)$ - funkce času nebo náhodná veličina)
- $E\{w\} = L_w$ – střední počet požadavků ve frontě (střední délka)
- t_w – doba čekání jednoho konkrétního požadavku (náhodná veličina)
- $E\{t_w\} = T_w$ - střední doba čekání požadavku ve frontě



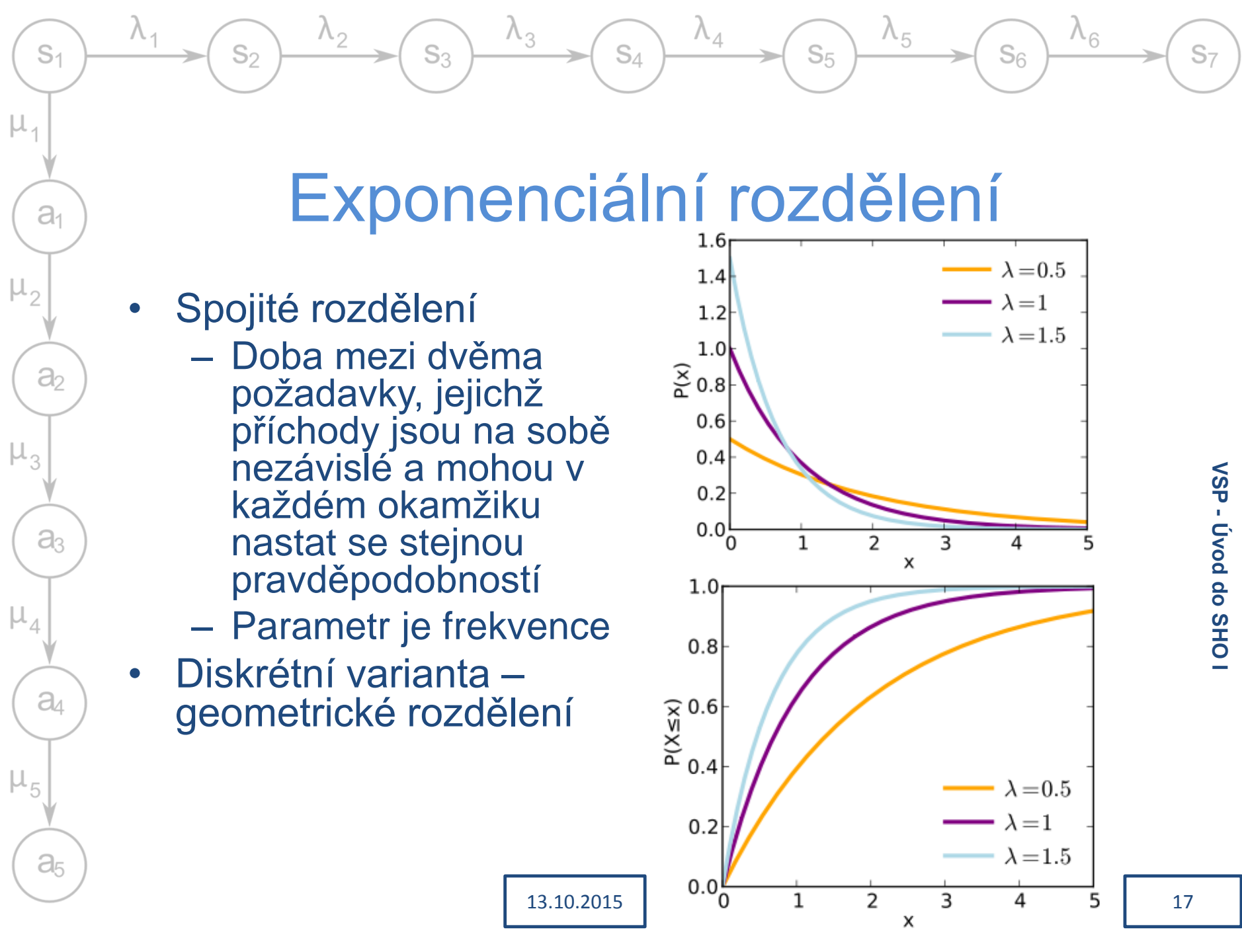
Kanály obsluhy

- Modelují skutečné obslužné prvky
 - Pro jedno elementární SHO jich může být obecně m (např. server se 4 jádry a jednou frontou), nejčastěji $m = 1$
- Doba obsluhy t_s chápána jako náhodná veličina
 - pro charakteristiku kanálu $F_s = P\{t_s < t\}$
 - V reálném případě lze odvodit aproximací z histogramu
 - Jak výkonnost kanálu tak specifika požadavků na dobu obsluh
 - Stejné pravděpodobnostní rozdělení pro všechny požadavky (jinak nejde spočítat)



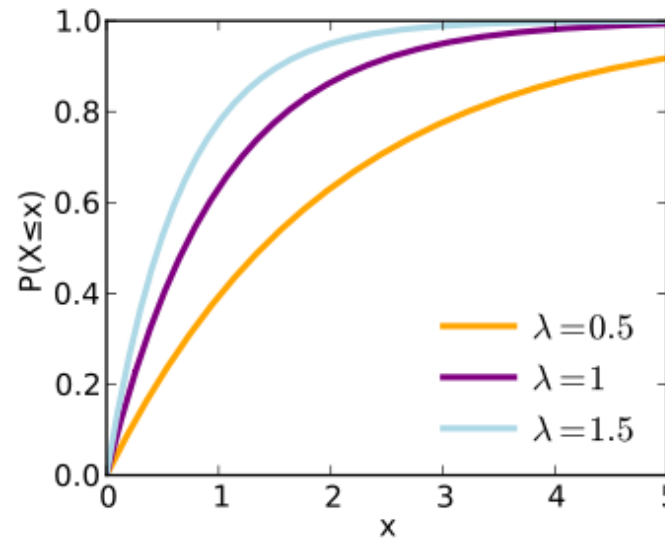
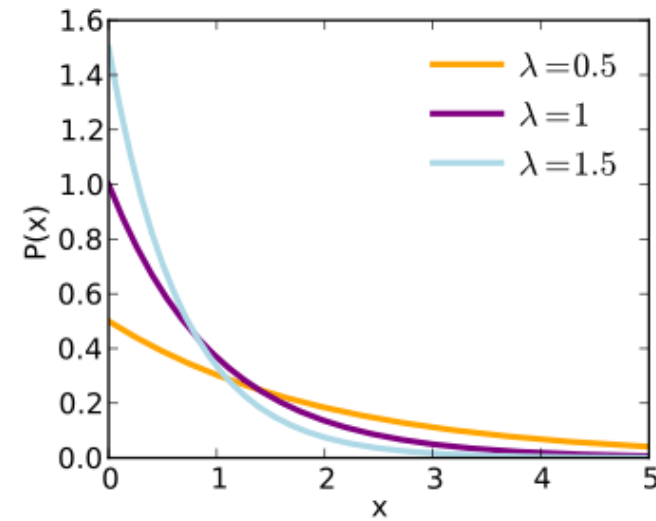
Charakteristika kanálu obsluhy

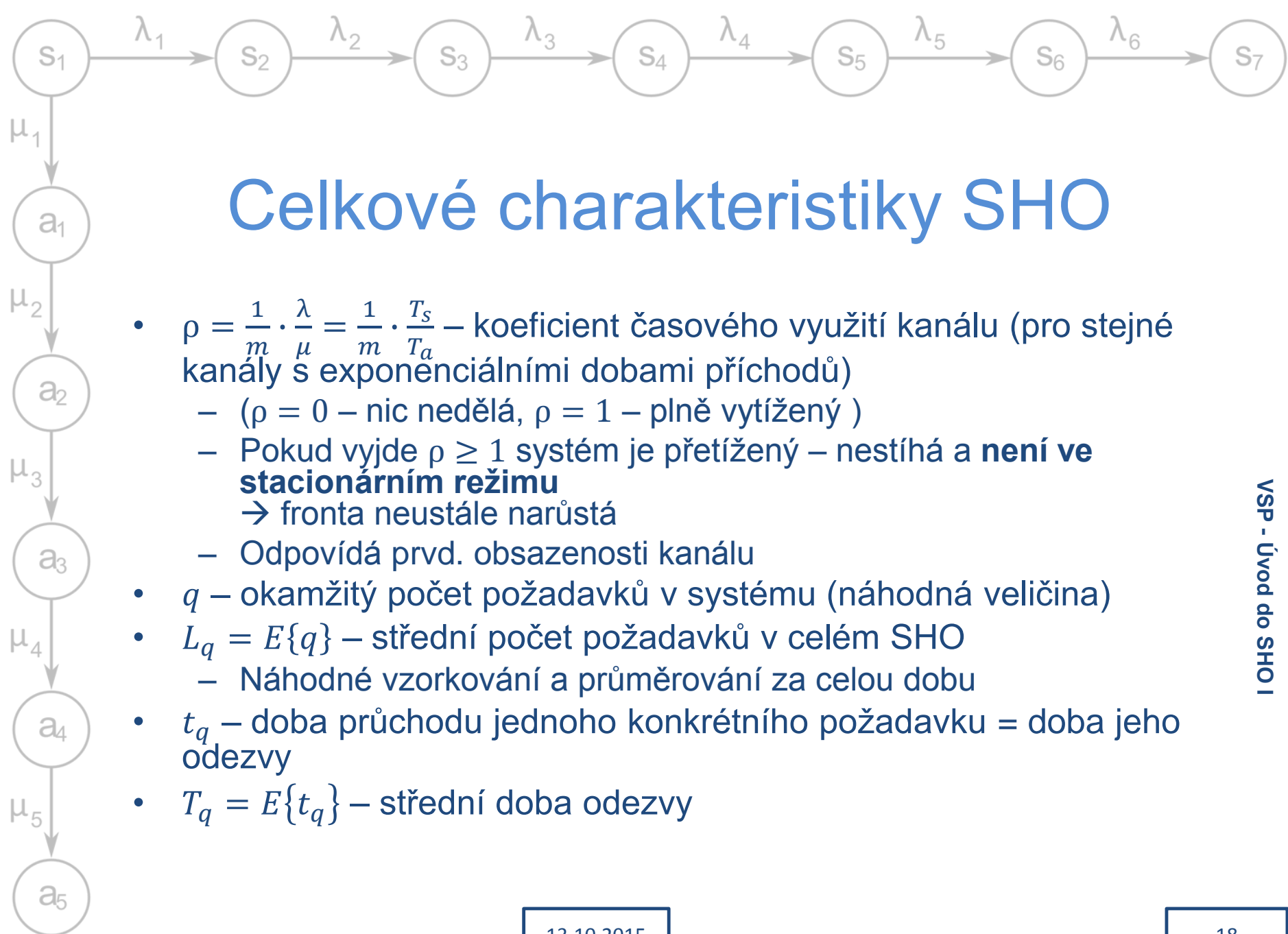
- Podobná jako charakteristika zdroje požadavků
- $F_S(t_S)$ – distribuční funkce pro doby obsluh
 - Pro exponenciální rozdělení $F_S(t_S) = 1 - e^{-\mu t}$, kde μ je střední frekvence obsluh (za předpokladu že kanál je stále zatížený!)
- $E\{\tau\} = T_S = \frac{1}{\mu}$ - střední doba obsluhy
- $C_S = \frac{\sigma\{t_S\}}{T_S}$ - koeficient variance
 „jak moc jsou doby obsluh náhodné“



Exponenciální rozdělení

- Spojité rozdělení
 - Doba mezi dvěma požadavky, jejichž příchody jsou na sobě nezávislé a mohou v každém okamžiku nastat se stejnou pravděpodobností
 - Parametr je frekvence
- Diskrétní varianta – geometrické rozdělení





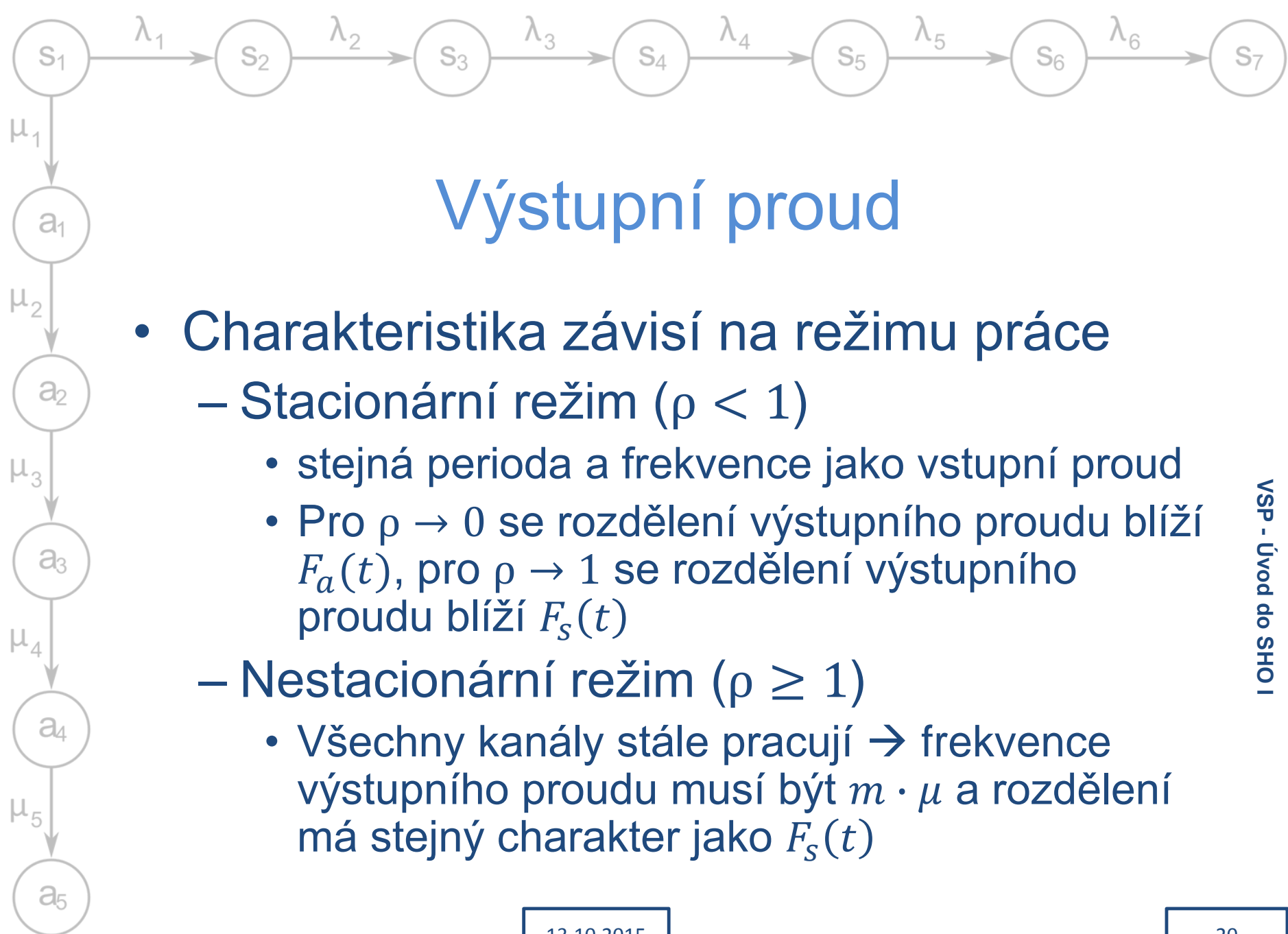
Celkové charakteristiky SHO

- $\rho = \frac{1}{m} \cdot \frac{\lambda}{\mu} = \frac{1}{m} \cdot \frac{T_s}{T_a}$ – koeficient časového využití kanálu (pro stejné kanály s exponenciálními dobami příchodů)
 - ($\rho = 0$ – nic nedělá, $\rho = 1$ – plně vytížený)
 - Pokud vyjde $\rho \geq 1$ systém je přetížený – nestíhá a **není ve stacionárním režimu**
 → fronta neustále narůstá
 - Odpovídá prvd. obsazenosti kanálu
- q – okamžitý počet požadavků v systému (náhodná veličina)
- $L_q = E\{q\}$ – střední počet požadavků v celém SHO
 - Náhodné vzorkování a průměrování za celou dobu
- t_q – doba průchodu jednoho konkrétního požadavku = doba jeho odezvy
- $T_q = E\{t_q\}$ – střední doba odezvy



Základní vztahy středních hodnot

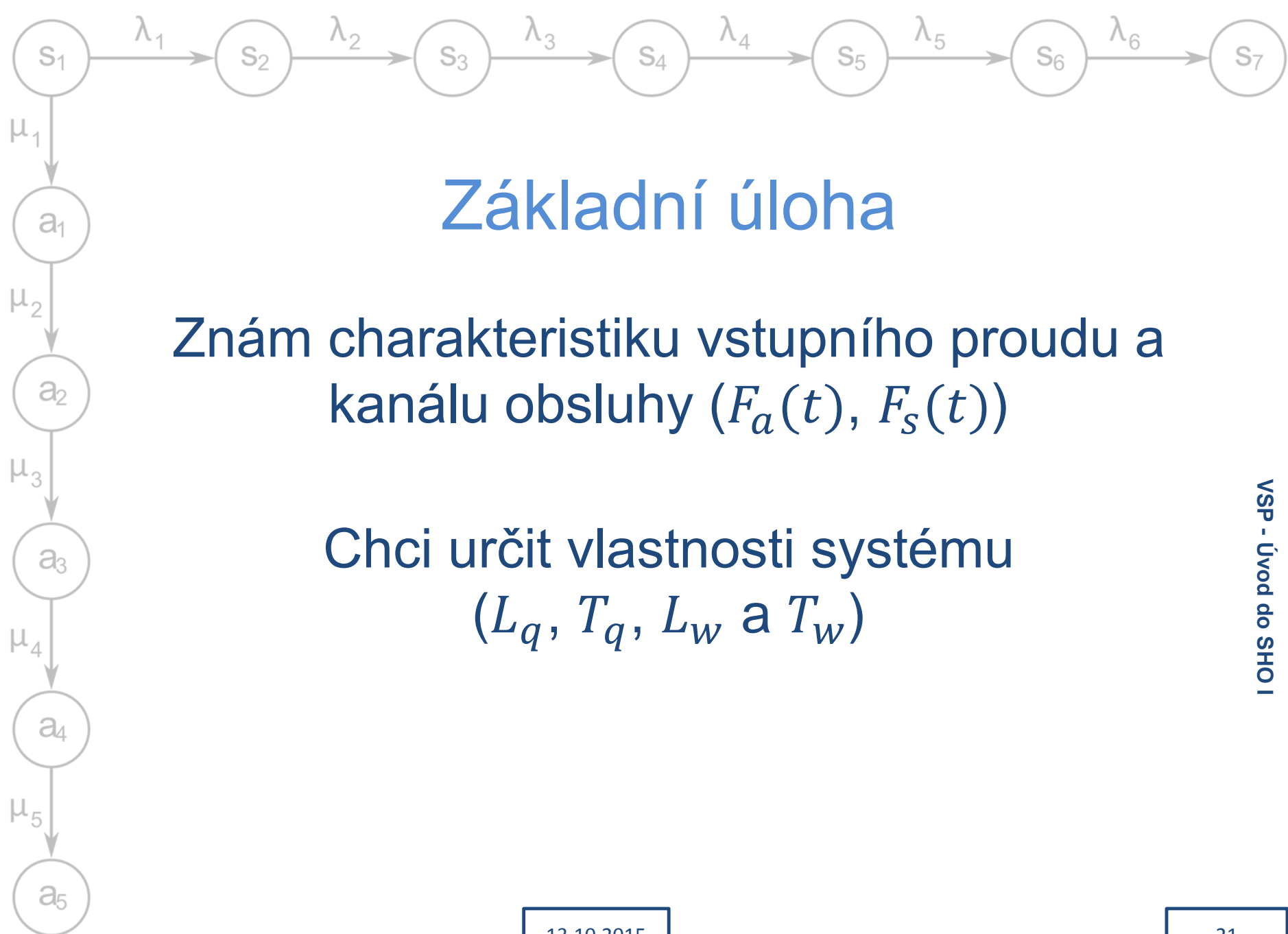
- $L_q = L_w + L_s = L_w + m \frac{\lambda}{\mu}$ - v systému je to co je ve frontě a v kanálech obsluhy
 - $T_q = T_w + T_s = T_w + \frac{1}{\mu}$ - průchod systémem je složen z čekání a obsluhy
 - **Littleovy vzorce** (stacionární režim, FIFO!)
 - $L_q = \lambda \cdot T_q$
 - $L_w = \lambda \cdot T_w$
 - $T_w = L_w \cdot T_a$
 - Nezávisí na pravděpodobnostním rozdělení vstupního proudu, lze uplatnit i na složený systém
- z jedné z hodnot L_q , T_q , L_w a T_w a znalosti λ lze dopočítat ostatní



Výstupní proud

- Charakteristika závisí na režimu práce
 - Stacionární režim ($\rho < 1$)
 - stejná perioda a frekvence jako vstupní proud
 - Pro $\rho \rightarrow 0$ se rozdělení výstupního proudu blíží $F_a(t)$, pro $\rho \rightarrow 1$ se rozdělení výstupního proudu blíží $F_s(t)$
 - Nestacionární režim ($\rho \geq 1$)
 - Všechny kanály stále pracují \rightarrow frekvence výstupního proudu musí být $m \cdot \mu$ a rozdělení má stejný charakter jako $F_s(t)$

VSP - Úvod do SHO I



Základní úloha

Znám charakteristiku vstupního proudu a kanálu obsluhy ($F_a(t)$, $F_s(t)$)

Chci určit vlastnosti systému (L_q , T_q , L_w a T_w)



Kendallova klasifikace

- Charakteristika různých typů SHO, podle vlastností vstupních proudů, obsluh, front ... (z 50. let)
- Systém je popsán trojicí (pěticí)

X / Y / m / I / disc.

- **X** – prvd. rozdělení vstupního proudu
- **Y** – prvd. rozdělení dob obsluh
- **m** – počet kanálů obsluhy
- **I** – maximální délka fronty (obvykle ∞ - pak se neuvádí)
- **disc.** – frontová disciplína (obvykle FIFO, pak se neuvádí)



Kendallova klasifikace

- Typické charakteristiky proudů:
 - GI – obecné náhodné rozdělení, intervaly příchodů statisticky nezávislé
 - G – obecné náhodné rozdělení
 - M – exponenciální rozdělení dob obsluh nebo příchodů požadavků (*Markovské*)
 - D – deterministické intervaly (konstantní nebo jinak pravidelné)
- Typické příklady: M/M/1, M/M/m, M/M/1/n/FIFO, M/G/1





M/M/1 (/∞/FIFO)



- Nejjednodušší varianta

- Poissonovský vstupní proud, exponenciální doby obsluhy → lze charakterizovat 2

parametry: $\frac{\lambda}{\mu}$

- λ – střední frekvence vstupního proudu
($F_a(t) = 1 - e^{-\lambda t}$)

- μ – střední podmíněná frekvence obsluh
pro $\rho < 1$ bude skutečná frekvence nižší

- $\rho = \frac{\lambda}{\mu} \rightarrow$ pro stacionární režim $\lambda < \mu$

- Lze modelovat jako markovský proces



Markovský model M/M/1



$$0 = -\lambda p_0 + \mu p_1$$

$$p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0$$

$$0 = \lambda p_0 - \lambda p_1 - \mu p_1 + \mu p_2$$

$$p_2 = \frac{\lambda}{\mu} p_1 = \rho p_1 = \rho^2 p_0$$

...

$$\rho_k = \rho^k p_0 = \rho^k (1 - \rho)$$

- Prvd. funkce rozložení náhodné veličiny k (počet požadavků v SHO)



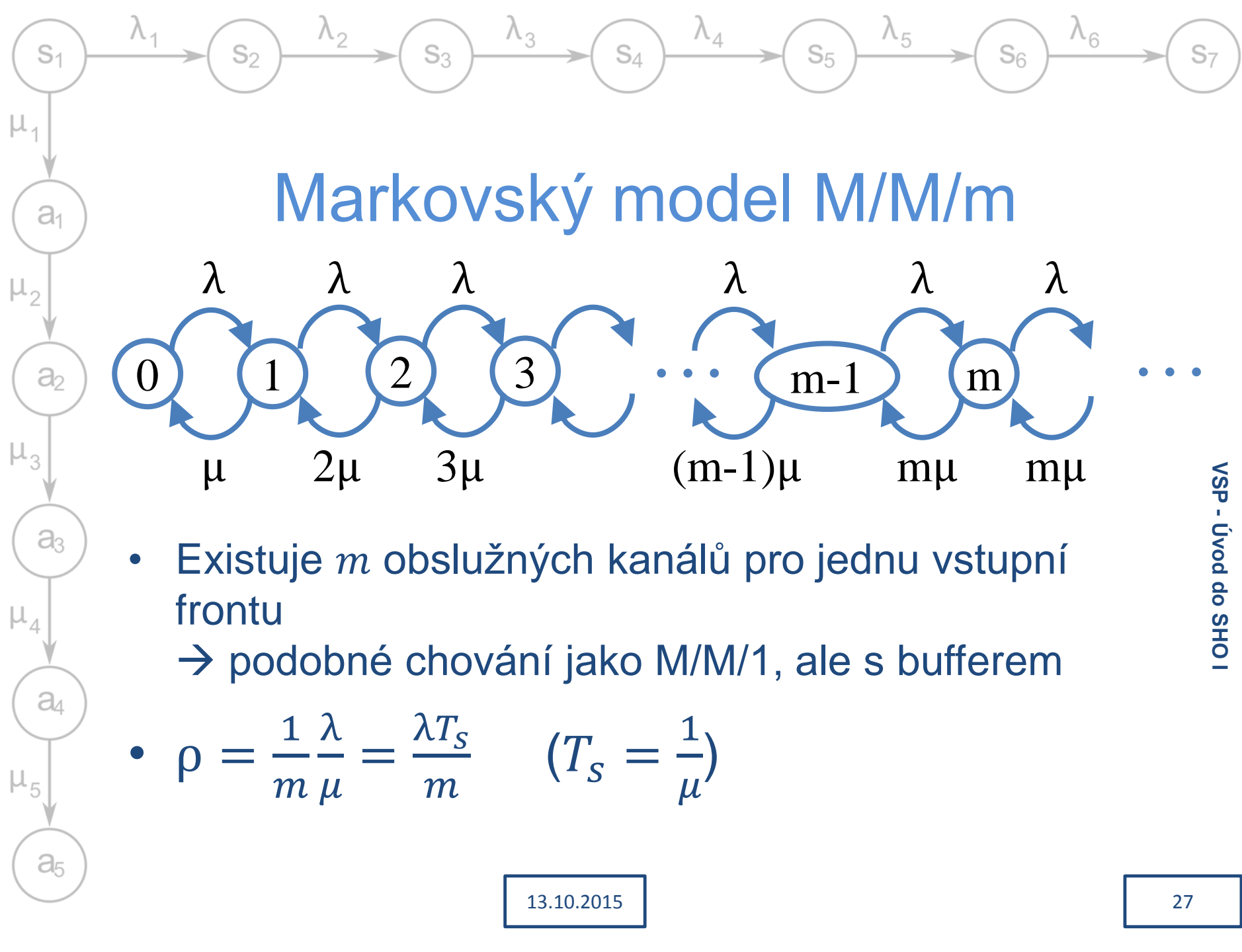
Markovský model M/M/1



$$\rho_k = \rho^k p_0 = \rho^k (1 - \rho)$$

- $$E\{k\} = \sum_{k=0}^{\infty} k p_k = (1 - \rho) \sum_{k=0}^{\infty} k \rho^k = (1 - \rho) \frac{\rho}{(1 - \rho)^2}$$

$$= \frac{\rho}{(1 - \rho)} = L_q \text{ (pro } (\rho < 1))$$
- T_q, L_w a T_w lze určit z Littleových vzorců



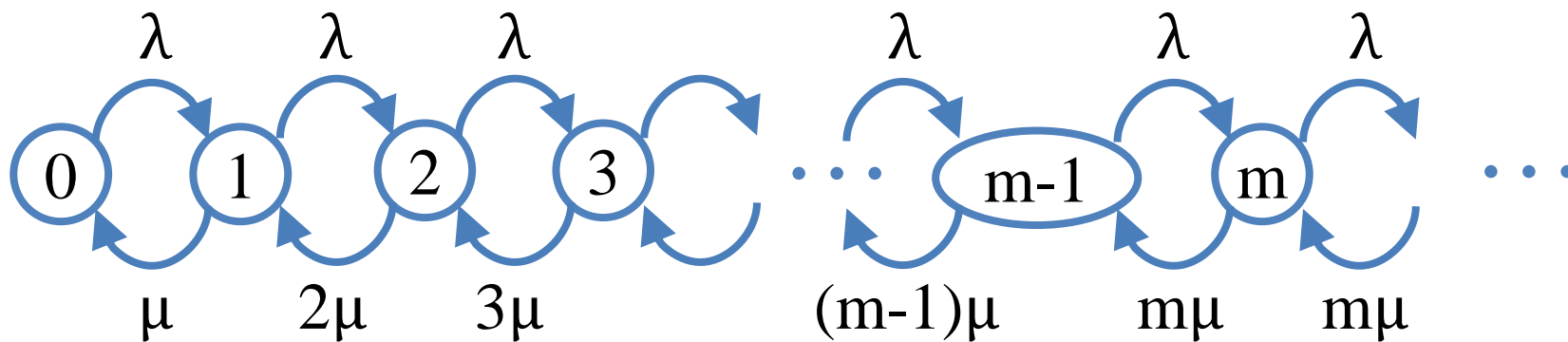
Markovský model M/M/m

- Existuje m obslužných kanálů pro jednu vstupní frontu
 → podobné chování jako M/M/1, ale s bufferem

- $$\rho = \frac{1}{m} \frac{\lambda}{\mu} = \frac{\lambda T_s}{m} \quad \left(T_s = \frac{1}{\mu} \right)$$



Markovský model M/M/m



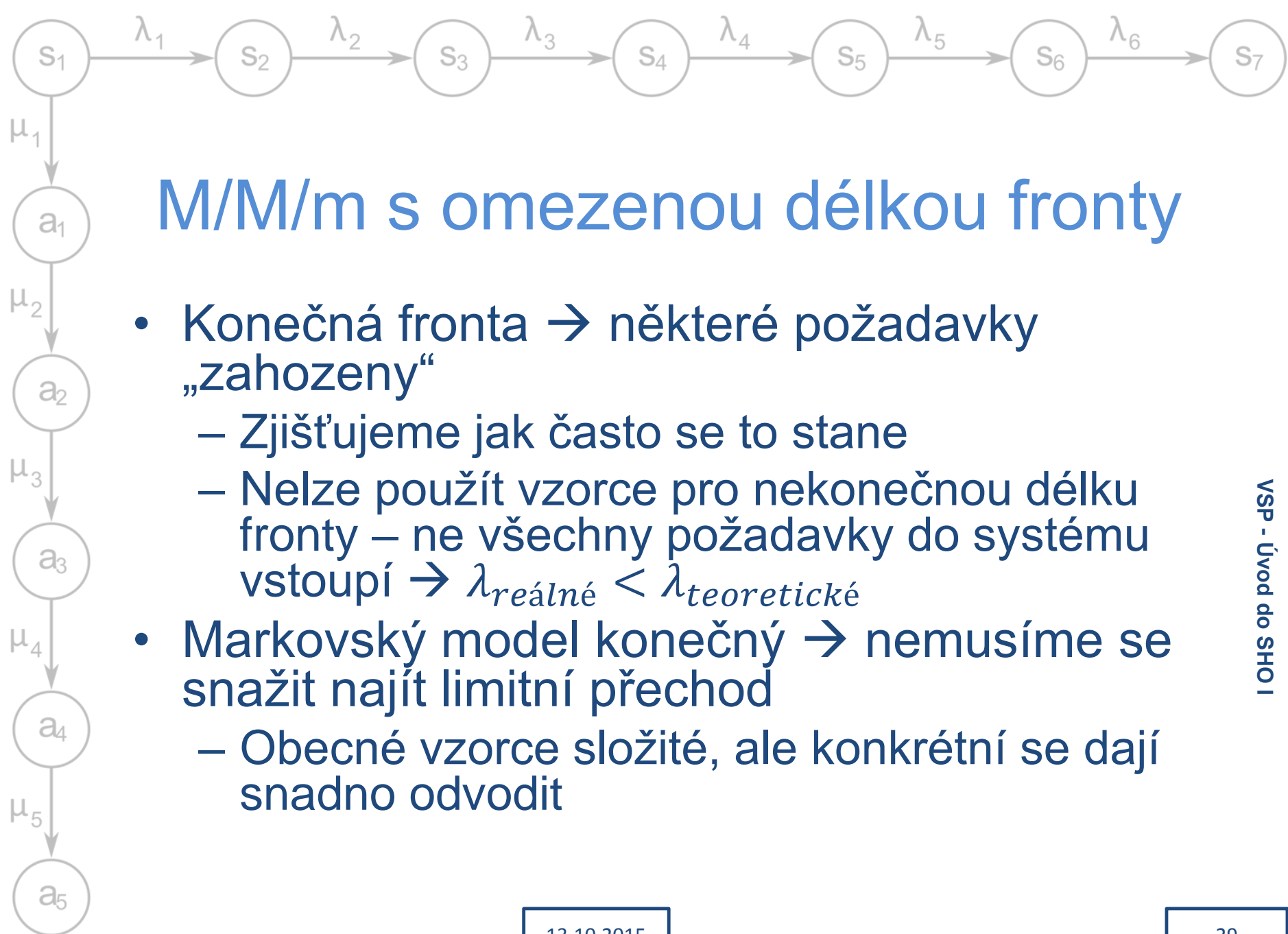
- Existuje obecné (složitě řešení), pro přibližný odhad (nebo $m \in \{1,2\}$):

$$- L_q \cong \frac{m\rho}{1-\rho^m}$$

$$- T_q \cong \frac{T_s}{1-\rho^m}$$

Řešení v **Probability, Statistics, and Queuing Theory** (Second Edition) nebo na

http://irh.inf.unideb.hu/~jsztrik/education/16/SOR_Main_Angol.pdf (<http://irh.inf.unideb.hu/user/jsztrik/>)



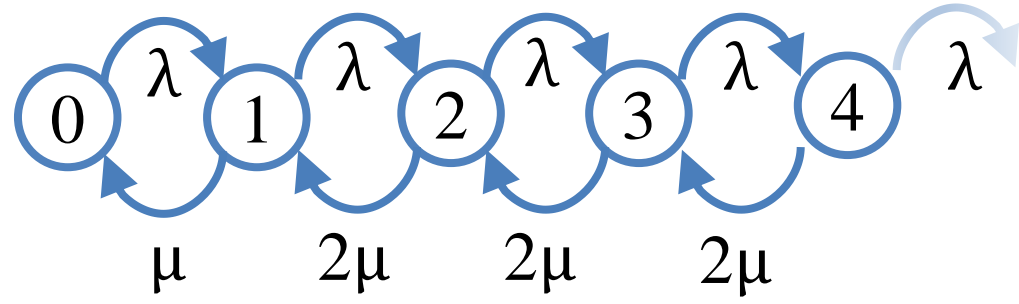
M/M/m s omezenou délkou fronty

- Konečná fronta → některé požadavky „zahozeny“
 - Zjišťujeme jak často se to stane
 - Nelze použít vzorce pro nekonečnou délku fronty – ne všechny požadavky do systému vstoupí → $\lambda_{reálné} < \lambda_{teoretické}$
- Markovský model konečný → nemusíme se snažit najít limitní přechod
 - Obecné vzorce složité, ale konkrétní se dají snadno odvodit

VSP - Úvod do SHO I



M/M/2/2/FIFO

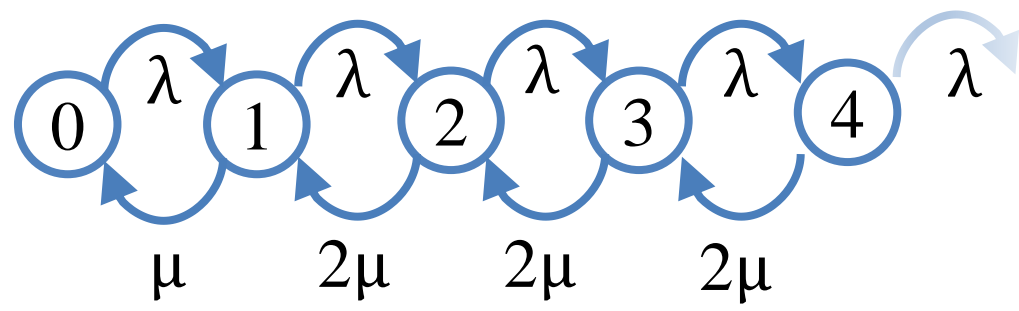


- 2 servery, fronta pro 2 požadavky

- Lze sestavit soustavu lineárních rovnic a najít p_0, p_1, p_2, p_3, p_4
- p_0 - pravděpodobnost že systém nic nedělá
- p_4 - pravděpodobnost že žádný další požadavek nemůže vstoupit ($p_4 \cdot 100\%$ - podíl odmítnutých požadavků)
- $L_q = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4$

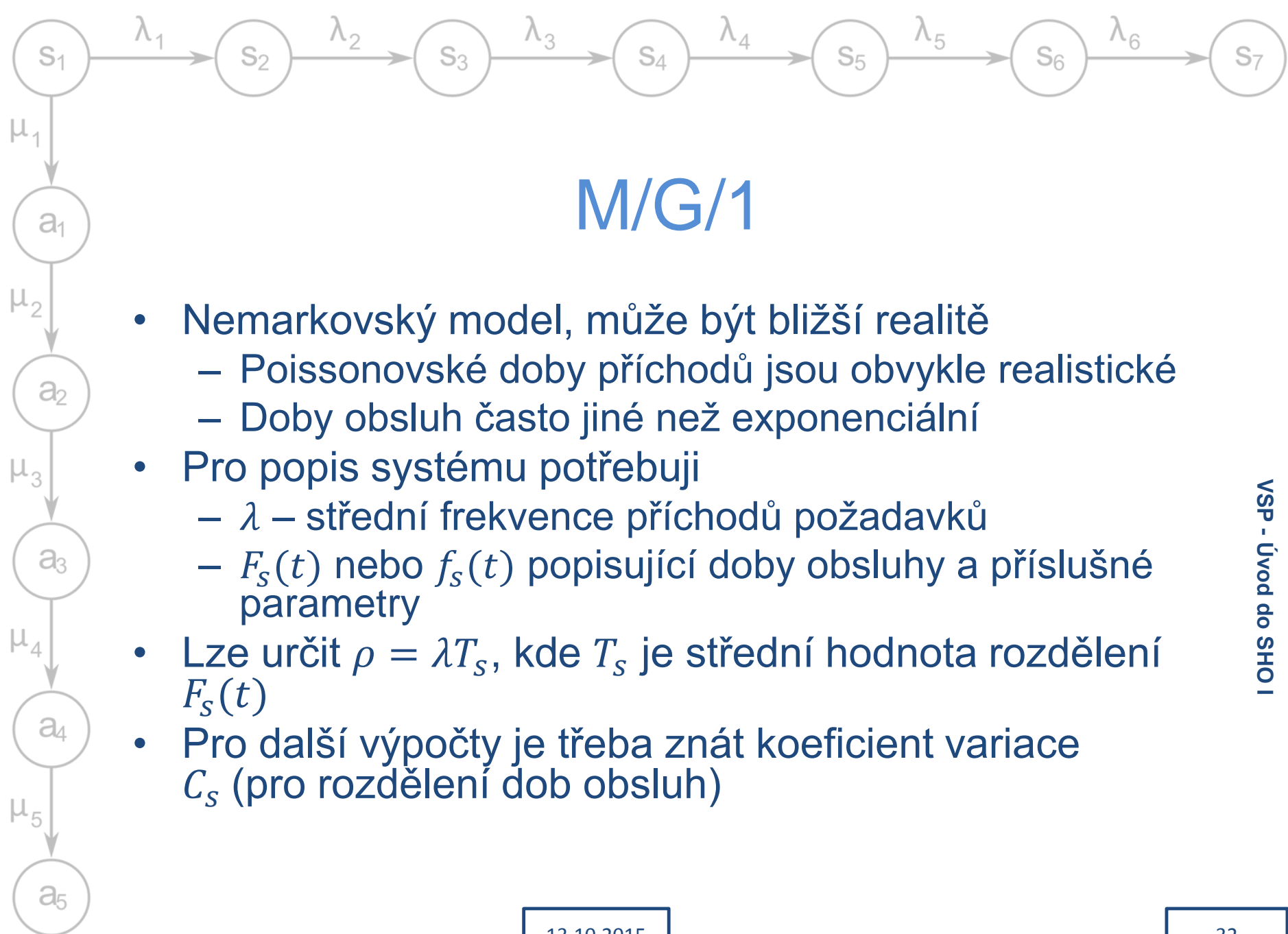


M/M/2/2/FIFO



- 2 servery, fronta pro 2 požadavky

- $\rho = \frac{p_1}{2} + p_2 + p_3 + p_4$
 - Vážený průměr zatížení v jednotlivých stavech
- $L_w = p_3 + 2p_4$
 - Vážený průměr možných délek front
- Střední frekvence zahození požadavku: λp_4
 - Frekvence přijetí: $\lambda(1 - p_4)$ - skutečná frekvence příchoďů požadavků
- $T_q = \frac{L_q}{\lambda(1-p_4)}$ (Little - $L_q = \lambda T_q$)



- Nemarkovský model, může být bližší realitě
 - Poissonovské doby příchodů jsou obvykle realistické
 - Doby obsluh často jiné než exponenciální
- Pro popis systému potřebují
 - λ – střední frekvence příchodů požadavků
 - $F_s(t)$ nebo $f_s(t)$ popisující doby obsluhy a příslušné parametry
- Lze určit $\rho = \lambda T_s$, kde T_s je střední hodnota rozdělení $F_s(t)$
- Pro další výpočty je třeba znát koeficient variace C_s (pro rozdělení dob obsluh)



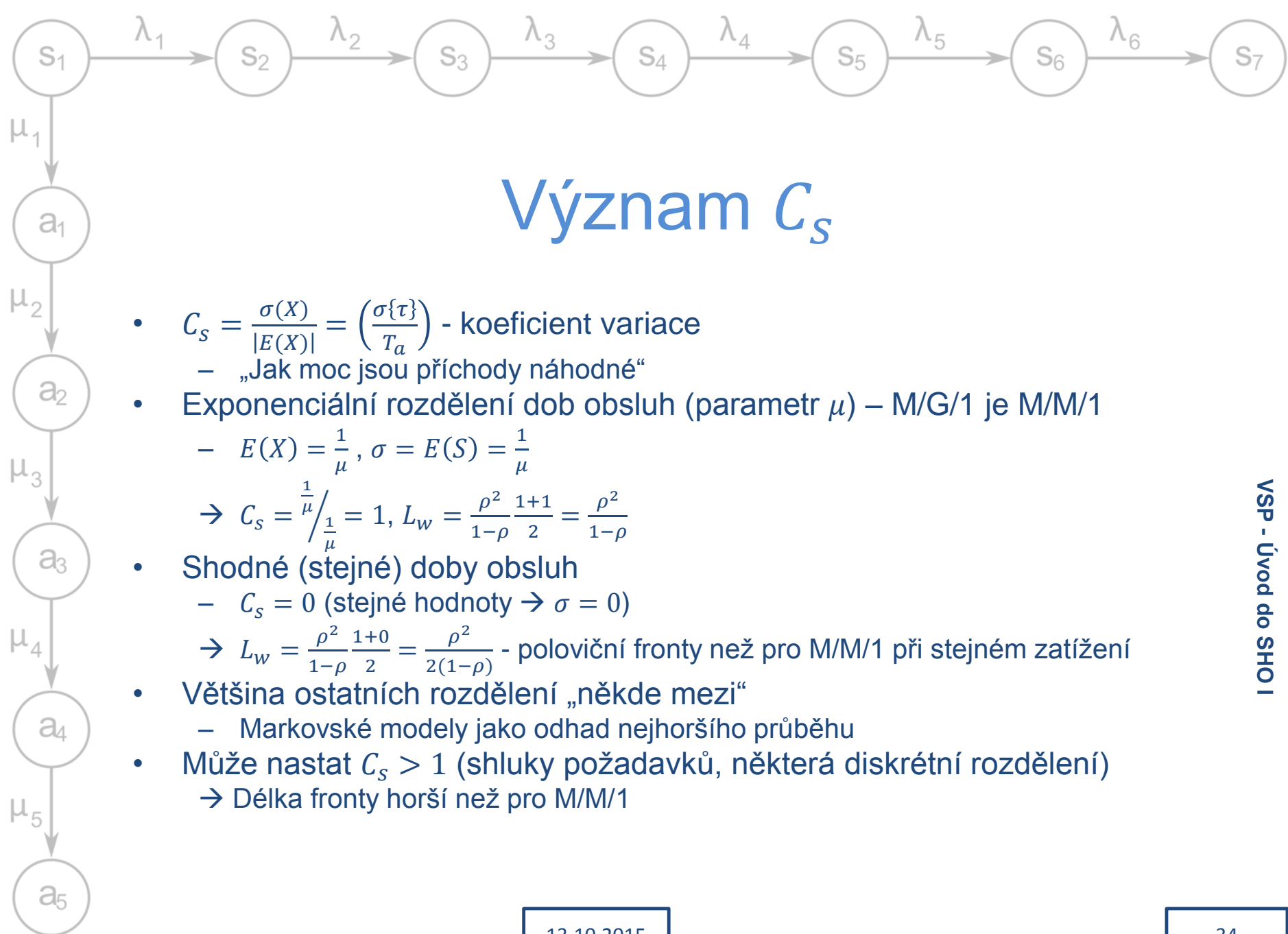
M/G/1

- $L_w = L_{w(M/M/1)} \frac{1+C_s^2}{2} = \frac{\rho^2}{1-\rho} \frac{1+C_s^2}{2}$ (střední délka fronty)

$$- L_q = L_w + L_s = L_w + \frac{\lambda}{\mu} = L_w + \rho$$

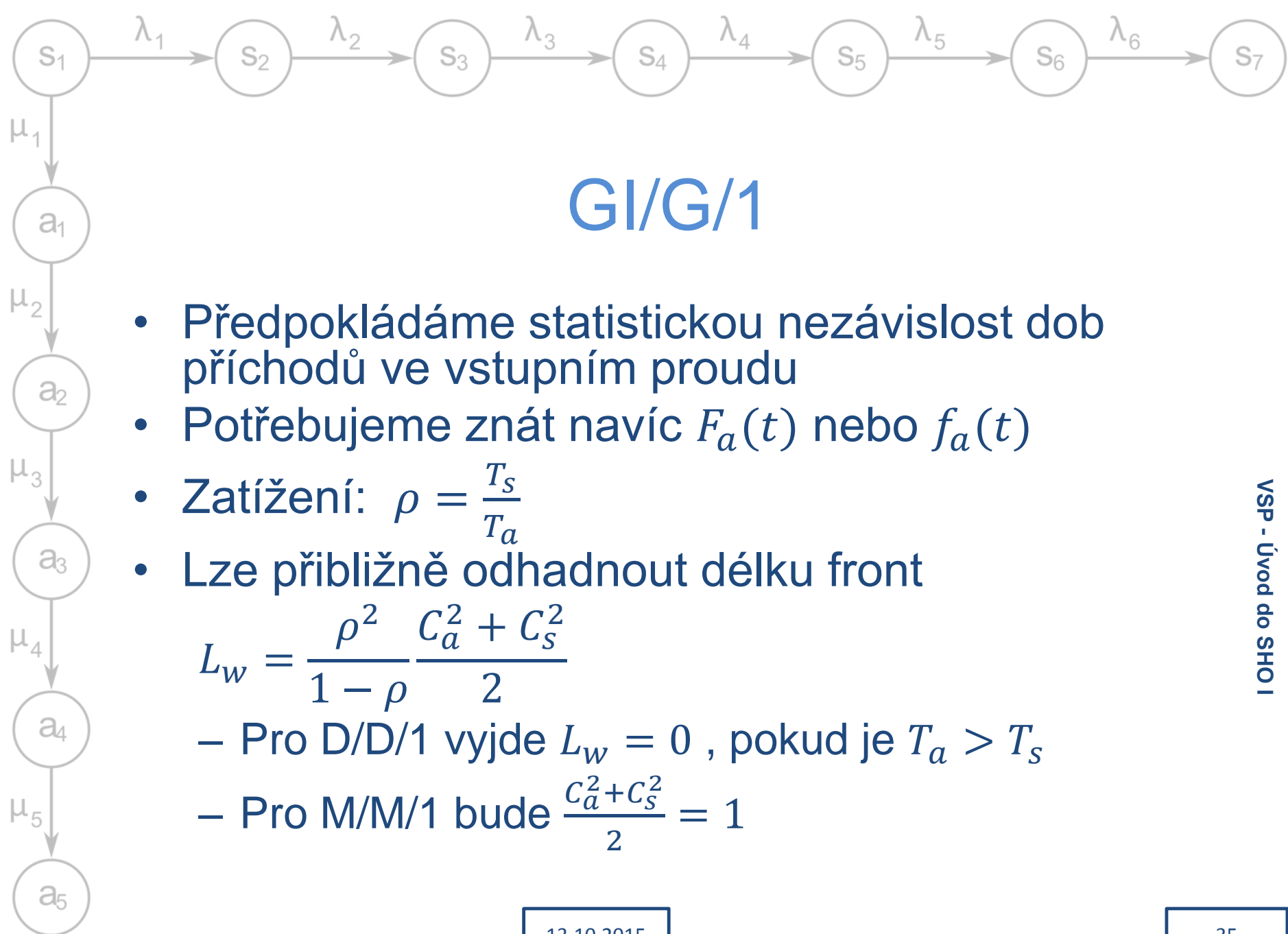
$$- T_q = \frac{L_q}{\lambda}$$

$$- T_w = \frac{L_w}{\lambda}, \quad T_w = T_q - T_s$$



Význam C_s

- $C_s = \frac{\sigma(X)}{|E(X)|} = \left(\frac{\sigma\{\tau\}}{T_a} \right)$ - koeficient variace
 - „Jak moc jsou příchody náhodné“
- Exponenciální rozdělení dob obsluh (parametr μ) – M/G/1 je M/M/1
 - $E(X) = \frac{1}{\mu}$, $\sigma = E(S) = \frac{1}{\mu}$
 - $C_s = \frac{\frac{1}{\mu}}{\frac{1}{\mu}} = 1$, $L_w = \frac{\rho^2}{1-\rho} \frac{1+1}{2} = \frac{\rho^2}{1-\rho}$
- Shodné (stejně) doby obsluh
 - $C_s = 0$ (stejně hodnoty → $\sigma = 0$)
 - $L_w = \frac{\rho^2}{1-\rho} \frac{1+0}{2} = \frac{\rho^2}{2(1-\rho)}$ - poloviční fronty než pro M/M/1 při stejném zatížení
- Většina ostatních rozdělení „někde mezi“
 - Markovské modely jako odhad nejhoršího průběhu
- Může nastat $C_s > 1$ (shluky požadavků, některá diskrétní rozdělení)
 - Délka fronty horší než pro M/M/1



GI/G/1

- Předpokládáme statistickou nezávislost dob příchodů ve vstupním proudu
- Potřebujeme znát navíc $F_a(t)$ nebo $f_a(t)$
- Zatížení: $\rho = \frac{T_s}{T_a}$
- Lze přibližně odhadnout délku front

$$L_w = \frac{\rho^2}{1 - \rho} \frac{C_a^2 + C_s^2}{2}$$

– Pro D/D/1 vyjde $L_w = 0$, pokud je $T_a > T_s$

– Pro M/M/1 bude $\frac{C_a^2 + C_s^2}{2} = 1$



Děkuji za pozornost

- Příště generátory náhodných čísel a metoda Monte Carlo

Úvod do SHO – síť front



Výkonnost a spolehlivost – KIV/VSP

Richard Lipka

20.10.2015



Malá odbočka k pravděpodobnosti

- Diskrétní rozdělení (každé pro které máme konkrétní hodnoty se chová jako diskrétní!!!)
 - Jednotlivé hodnoty, každá má nenulovou pravděpodobnost
 - $E\{X\} = \frac{1}{n} \sum_{k=1}^n x_k$, ale existuje jen několik (i) různých x_k (tříd), každé s četností m_k a pravděpodobnost výskytu x_k je tedy $\frac{m_k}{n} = p_k \rightarrow E\{X\} = \sum_{k=0}^i x_k p_k = \sum_{k=1}^i x_k \frac{m_k}{n}$
- Příklad „vážený vs. normální průměr“:
 - $X = \{1, 1, 1, 2, 2, 3, 3, 3\}$ - tři třídy, $\{1, 2, 3\}$
 - $p_1 = \frac{3}{8}, p_2 = \frac{2}{8}, p_3 = \frac{3}{8}$
 - $E\{X\} = \frac{1}{8} (1 + 1 + 1 + 2 + 2 + 3 + 3 + 3) = \frac{1}{8} (1 + 1 + 1) + \frac{1}{8} (2 + 2) + \frac{1}{8} (3 + 3 + 3) = 1 \cdot \frac{3}{8} + 2 \cdot \frac{2}{8} + 3 \cdot \frac{3}{8}$



Malá odbočka k pravděpodobnosti

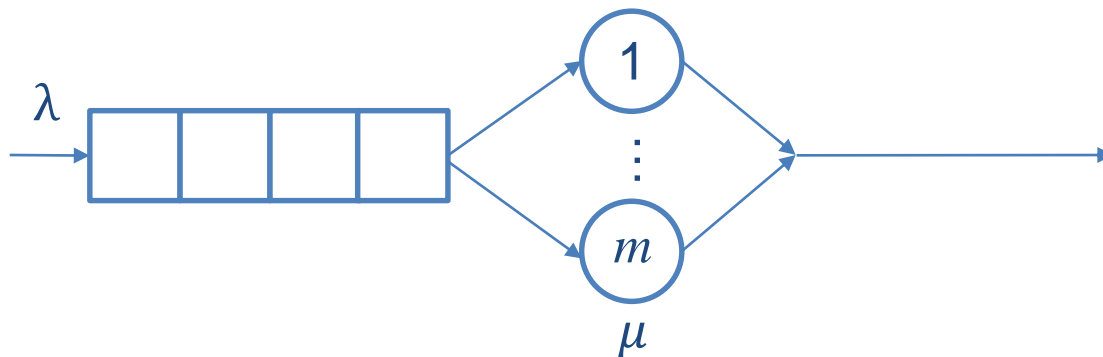
- Spojité rozdělení (teoretické, ne konkrétní hodnoty)
 - Nekonečně mnoho různých hodnot x_k , m výskytů konkrétní hodnoty x_k
 $\rightarrow p_k = \frac{m}{\infty} \rightarrow 0$, při konkrétní realizaci typicky $m \rightarrow 1$ (nebo velmi malé)
 - Pravděpodobnost lze určit jen pro jednotlivé intervaly: $p_{k,k+1} = \int_k^{k+1} f(x) dx$
- Střední hodnota definována:

$$E\{X\} = \int_{-\infty}^{\infty} x f(x) dx$$

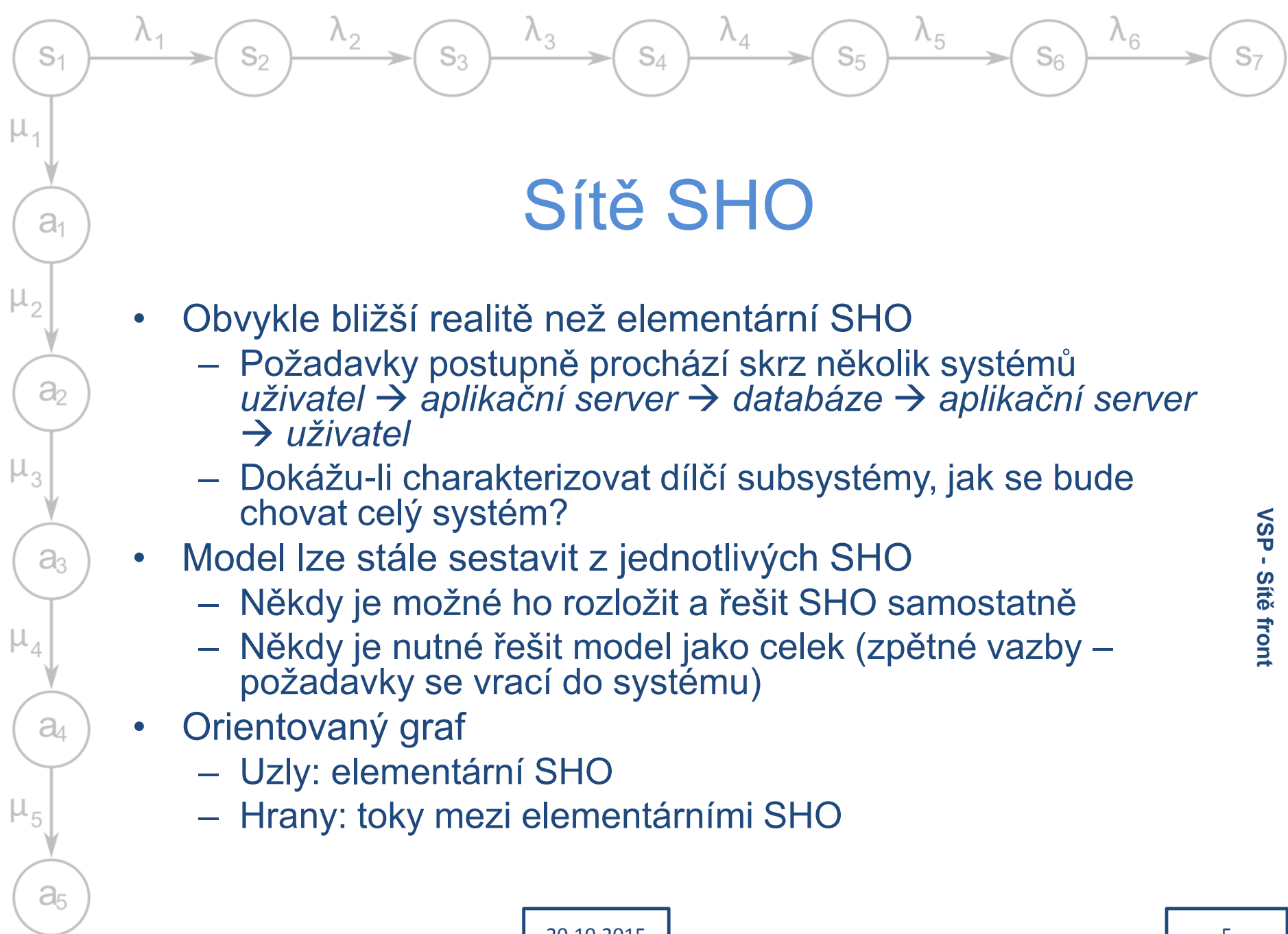
- x – konkrétní hodnota, $f(x)$ - „pravděpodobnost této hodnoty“
 \rightarrow podobné jako $E\{X\} = \sum_{k=0}^i x_k p_k$ ve spojitém světě



Elementární SHO - opakování

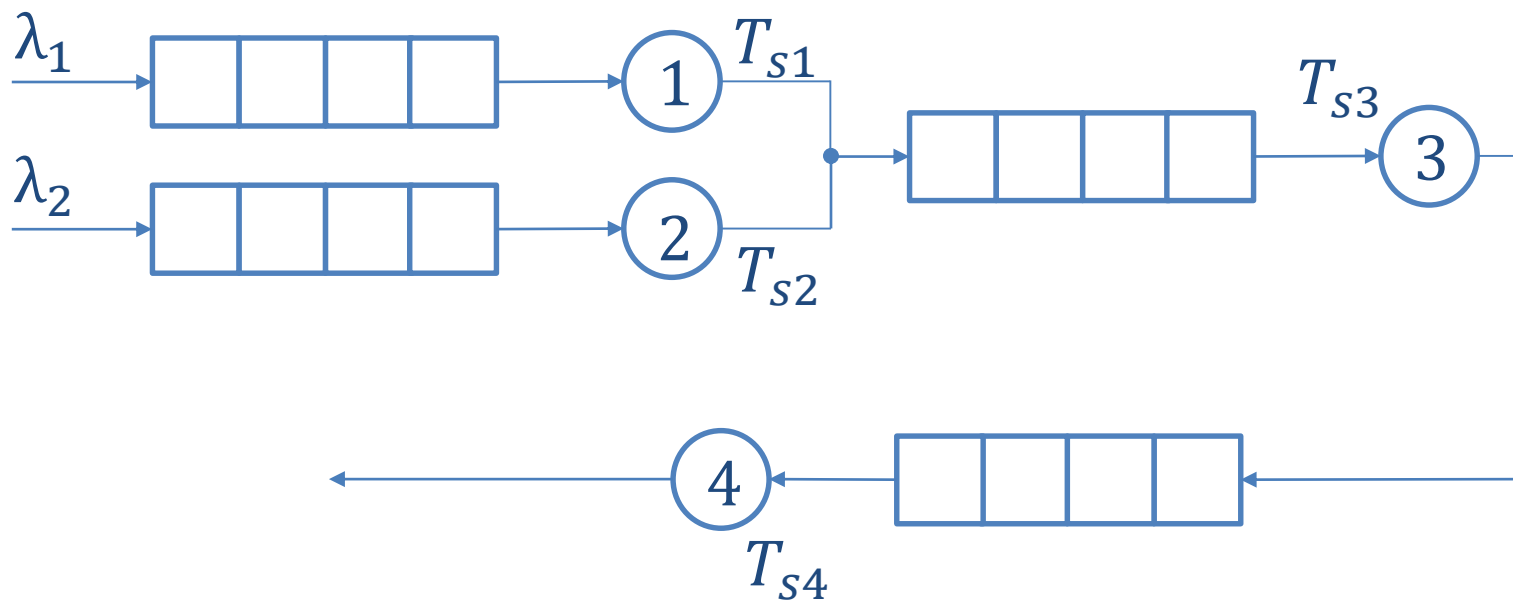


- Koeficient zatížení $\rho = \frac{1}{m} \cdot \frac{\lambda}{\mu} = \frac{1}{m} \cdot \frac{T_s}{T_a}$
 - Stacionární režim: $\rho < 1$
- Počet požadavků v systému $L_q \cong \frac{m\rho}{1-\rho^m}$
 - Pro $m \in \{1, 2\}$ přesné, jinak odhad
- Doba odezvy (obsluhy) $T_q \cong \frac{T_s}{1-\rho^m}$
- Littleovy vztahy: $L_q = \lambda \cdot T_q, L_w = \lambda \cdot T_w, T_w = L_w \cdot T_a$



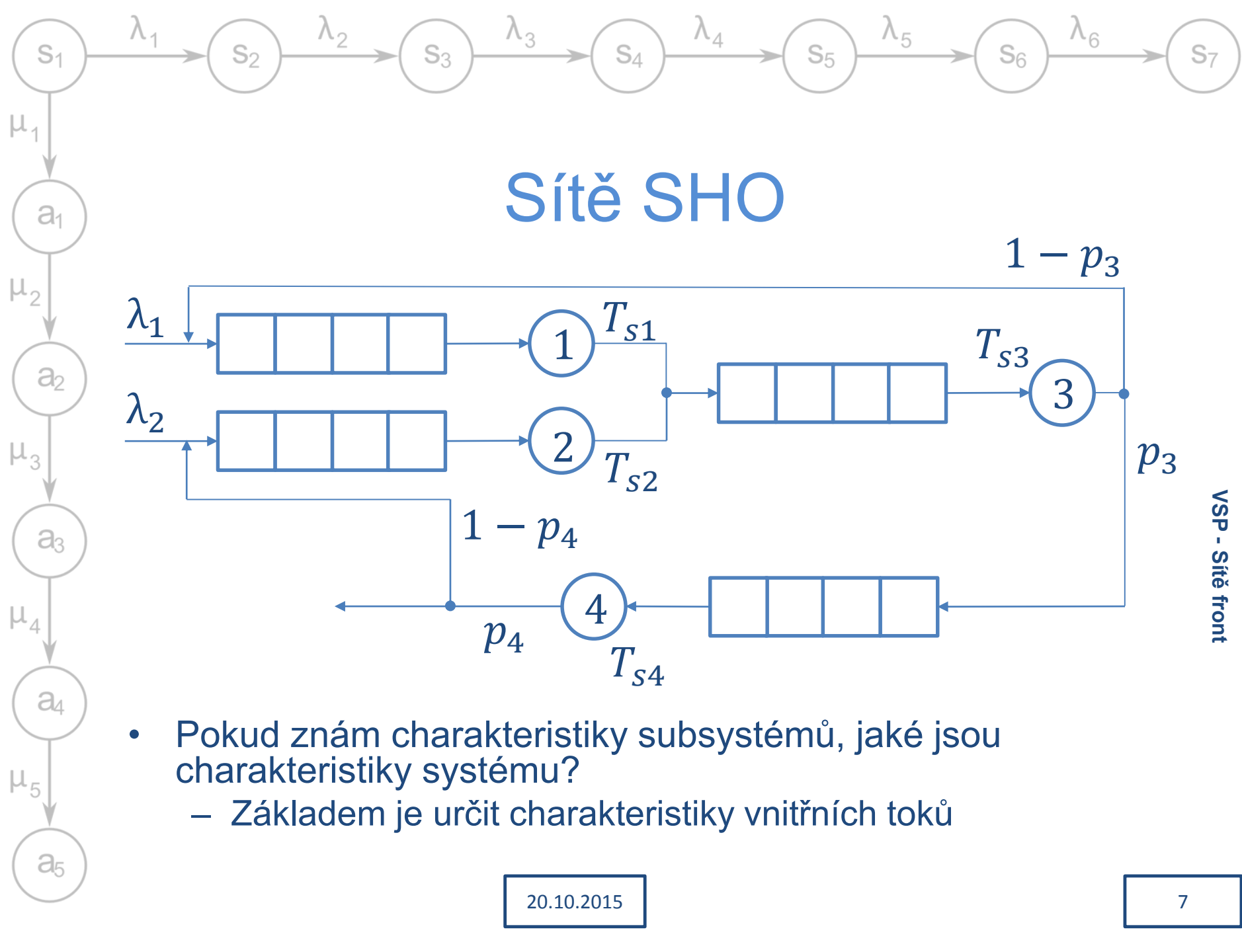


Sítě SHO



VSP - Síť front

- Jednotlivé systémy lze řešit odděleně
 - Do uzlu 3 vstupuje součet toků 1 a 2 – lze zjistit jejich analýzou
 - Do uzlu 4 vstupuje výstupní tok uzlu 3

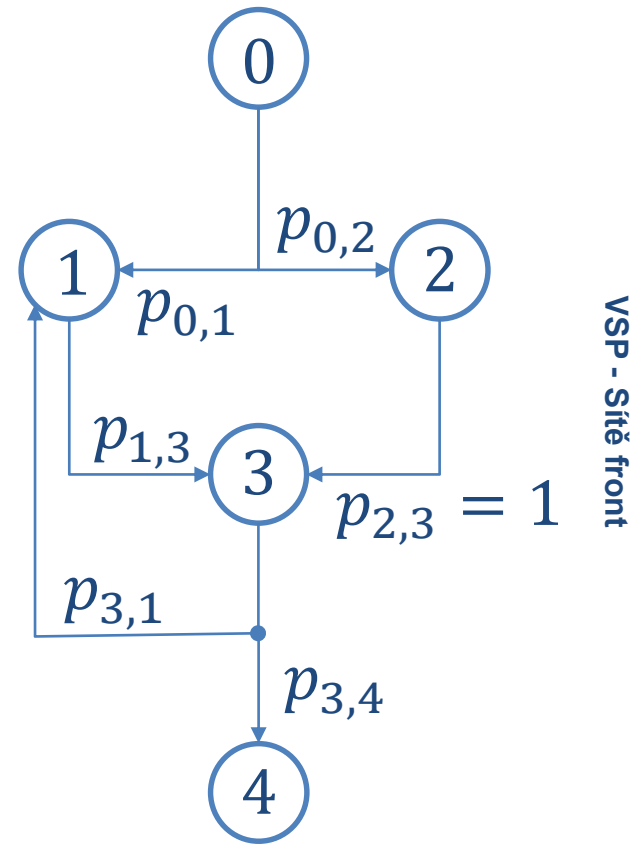




Popis SHO

Orientovaný graf

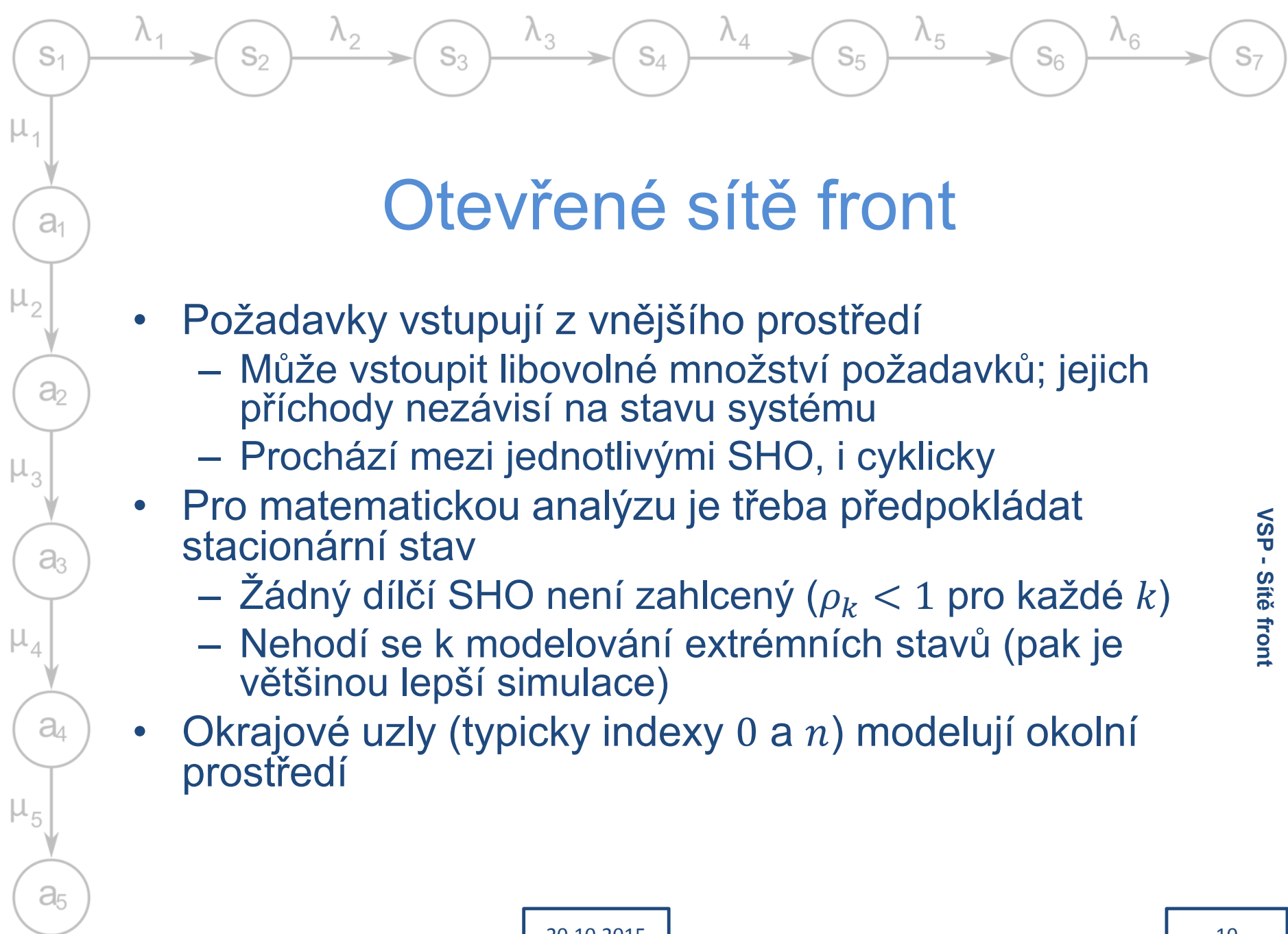
- Uzly – elementární SHO, n uzlů
- Hrany – toky mezi nimi
(→ hrana vede z výstupu obslužného kanálu do fronty jiného kanálu)
 - Intenzita toku do uzlu Λ_i - součet dílčích toků do i
 - Váha hrany: pravděpodobnost přechodu z uzlu i do uzlu j - $p_{i,j}$
- Okolí lze modelovat dalšími uzly (zdroj a stok požadavků)
- Zjevně platí $p_{3,1} + p_{3,4} = 1$
 - $\sum_{i=0}^n p_{k,n} = 1$ pro každý uzel k





Obecné vlastnosti sítí SHO

- Zákon zachování toků
 - „co do uzlu nebo křížení vteče, musí vytéct“
 - $\sum_{i=0}^n p_{k,n} = 1$ pro každý uzel k
 - Neplatí pro okrajové uzly v otevřených sítích – okolní prostředí
 - Spojování toků zachovává exponenciální rozdělení
- Výstupní tok z M/M/m
 - Stejná frekvence a intenzita jako u vstupního toku, charakter záleží na zatížení (očekáváme $\rho < 1$)
- „Jacksonův zákon“
 - Jsou-li všechny vstupní toky poissonovské a obsluhy exponenciální, jsou i výstupní toky exponenciální
 → jednotlivé dílčí SHO lze řešit jako M/M/m (pokud dokážeme určit vnitřní toky)

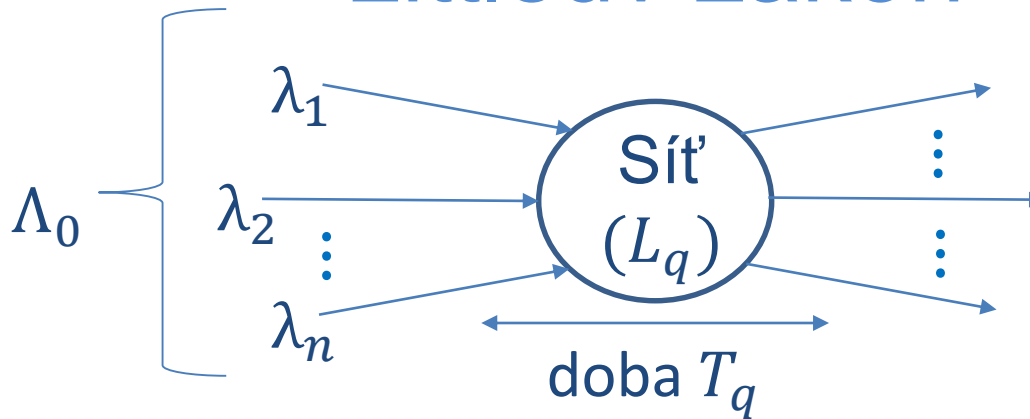


Otevřené sítě front

- Požadavky vstupují z vnějšího prostředí
 - Může vstoupit libovolné množství požadavků; jejich příchody nezávisí na stavu systému
 - Prochází mezi jednotlivými SHO, i cyklicky
- Pro matematickou analýzu je třeba předpokládat stacionární stav
 - Žádný dílčí SHO není zahlcený ($\rho_k < 1$ pro každé k)
 - Nehodí se k modelování extrémních stavů (pak je většinou lepší simulace)
- Okrajové uzly (typicky indexy 0 a n) modelují okolní prostředí



Littleův zákon



- $L_q = \Lambda_0 \cdot T_q$
 - počet požadavků v systému je úměrný intenzitě vstupů a době sledování
 - Stejné důvody jako u elementárních SHO (pokud je ve stacionárním stavu – nutno kontrolovat pro každý uzel zvlášť)

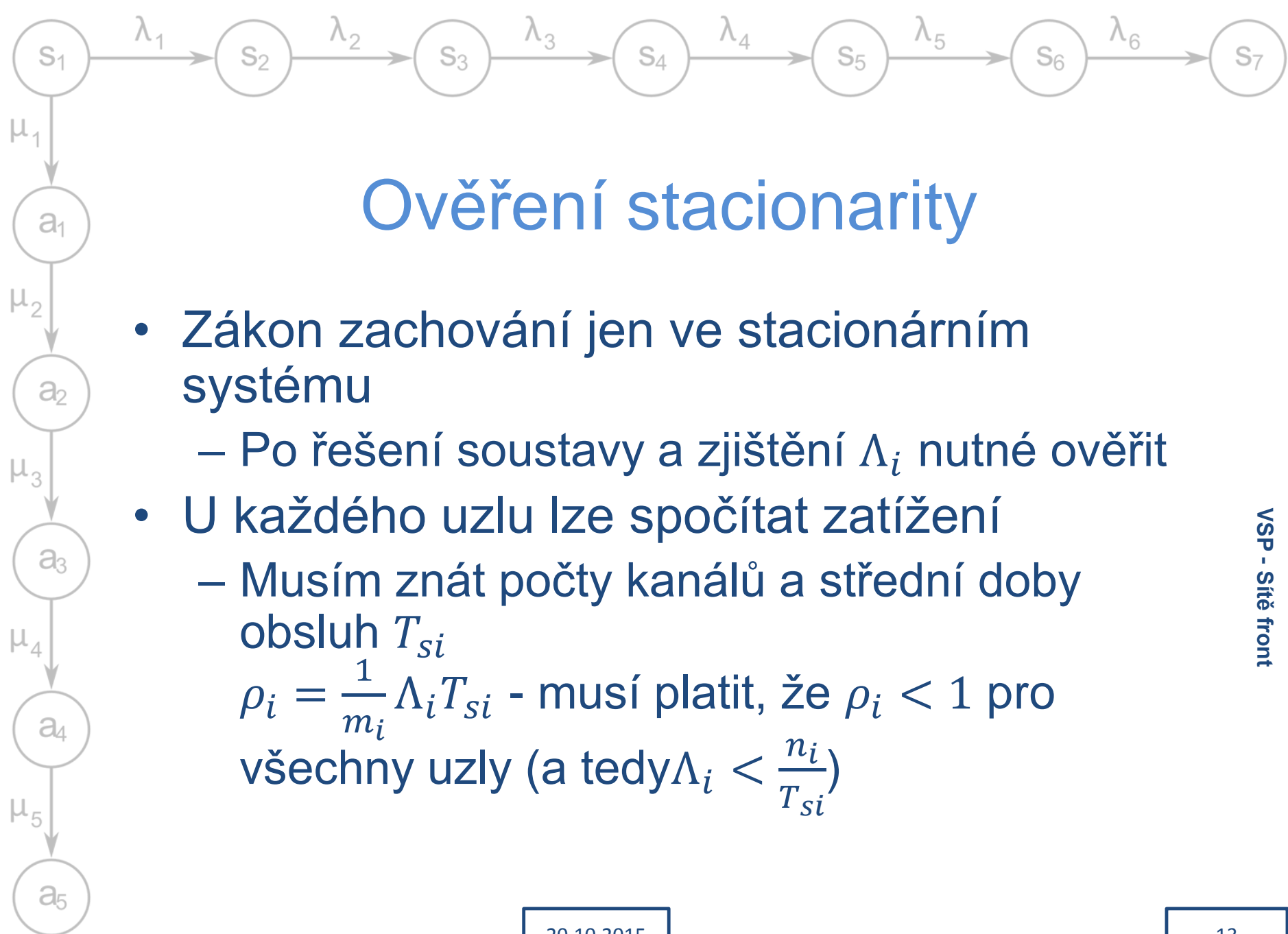


Střední frekvence a zatížení

VSP - Síť front

- Nezávisí na rozdělení, jen na středních frekvencích
- Λ_i - Vnitřní frekvence toku v uzlu
 - Λ_0 - souhrnný vstupní tok do systému
- Tok požadavků z i do j : $\lambda_{i,j} = \Lambda_i p_{i,j}$
 - Na vstupu uzlu i se sčítají frekvence na přicházejících hranách
- Stacionární režim a zákon zachování \rightarrow

$$\sum_k \Lambda_k p_{k,i} = \Lambda_i = \sum_j \Lambda_j p_{i,j}$$
 - k – uzly ze kterých vede hrana do i
 - „všechno co do uzlu vtéká z něj vytéká a nic navíc“
 - Řešením soustavy n rovnic získám vnitřní toky Λ_i



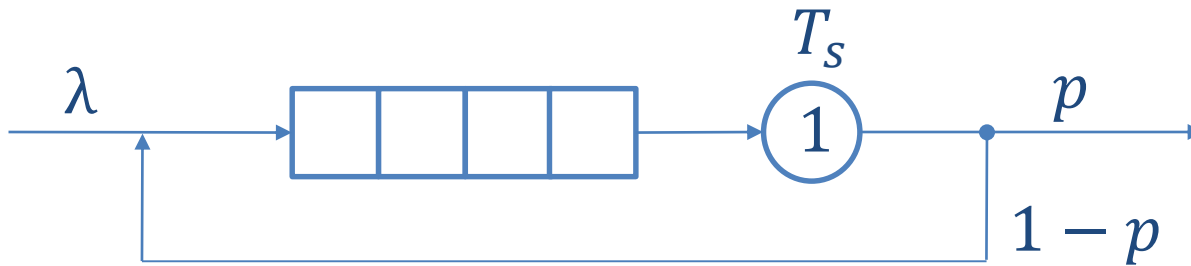
Ověření stacionarity

- Zákon zachování jen ve stacionárním systému
 - Po řešení soustavy a zjištění Λ_i nutné ověřit
- U každého uzlu lze spočítat zatížení
 - Musím znát počty kanálů a střední doby obsluh T_{si}
$$\rho_i = \frac{1}{m_i} \Lambda_i T_{si}$$

- musí platit, že $\rho_i < 1$ pro všechny uzly (a tedy $\Lambda_i < \frac{n_i}{T_{si}}$)



Příklad – finanční úřad



- Přichází 15 lidí za hodinu, obsluha trvá průměrně 3 minuty
 $\rightarrow \lambda = 15 \text{ hod}^{-1}, T_s = 0,05 \text{ hod}$
- Průměrně 1 formulář z 10 vyplněn špatně
 $\rightarrow p = 0,9$

$$\Lambda_1 = \lambda + \Lambda_1(1 - p)$$

$$\Lambda_1 = \frac{\lambda}{p} \cong 16,66$$

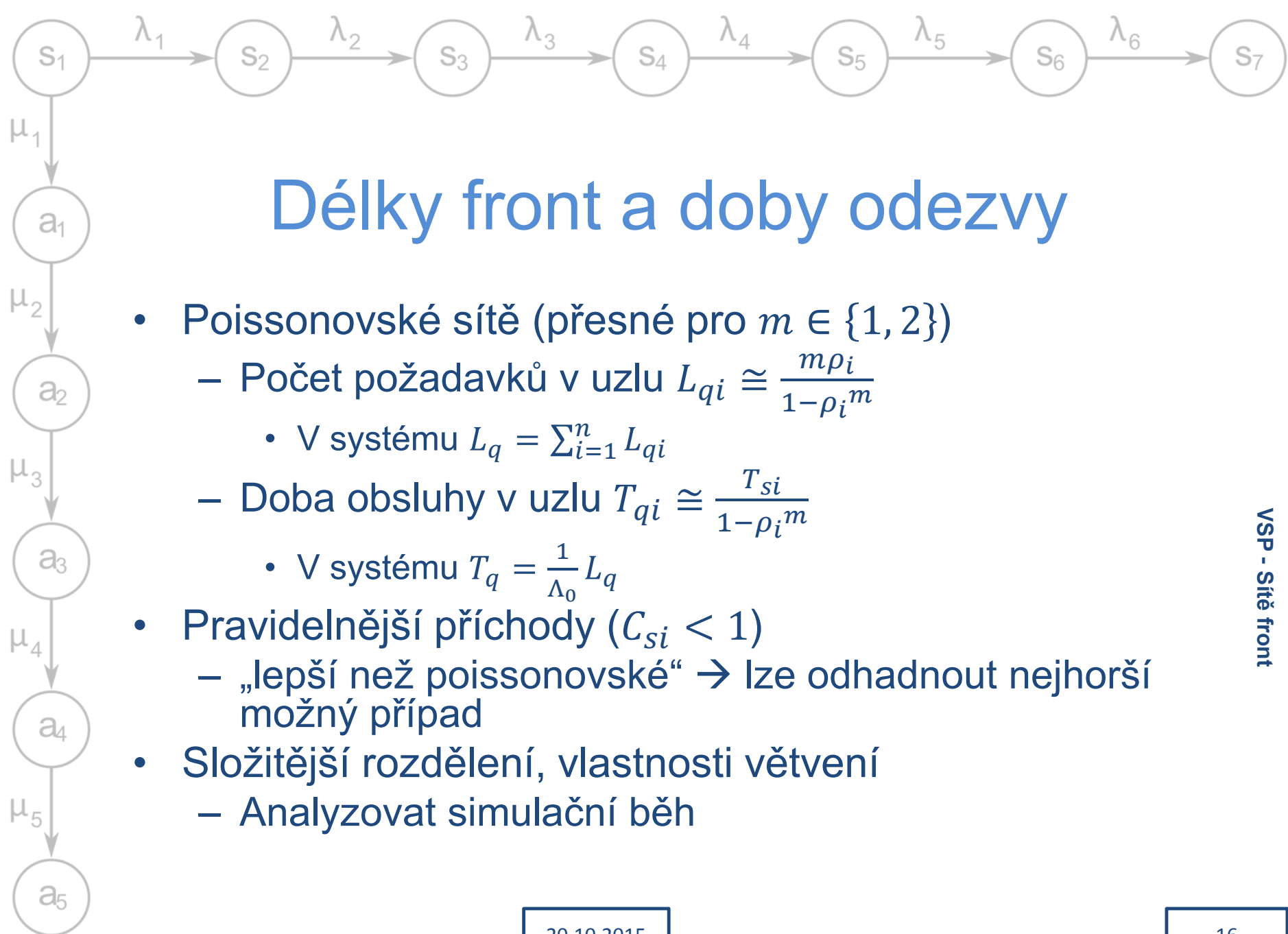
$$\rho_1 = \Lambda_1 T_s \cong 0,833$$

stacionární - OK



Délky front a doby odezvy

- Souvisí s charakterem rozdělení
 - Délky front narůstají s nepravidelností intervalů (synchronizované systémy fronty nevytváří)
 - Pro analytické řešení je třeba splnit Jacksonovy podmínky
 - Toky požadavků z okolí mají poissonovský charakter
 - Všechny doby obsluhy mají exponenciální rozdělení
 - Přejít mezi elementárními SHO je bez zpoždění, při větvení je náhodný, s pravděpodobností $p_{i,j}$
- výsledné toky budou také poissonovské a jednotlivé uzly (elementární SHO) lze analyzovat odděleně
- Frekvence vstupního toku Λ_i , střední doba obsluhy T_{si}

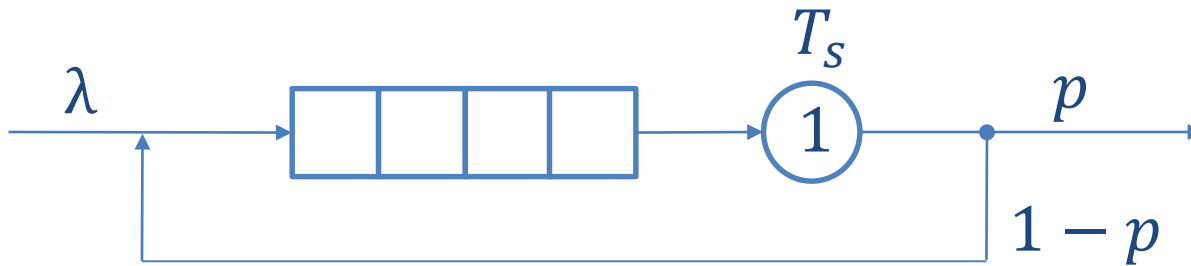


Délky front a doby odezvy

- Poissonovské sítě (přesné pro $m \in \{1, 2\}$)
 - Počet požadavků v uzlu $L_{qi} \cong \frac{m\rho_i}{1-\rho_i^m}$
 - V systému $L_q = \sum_{i=1}^n L_{qi}$
 - Doba obsluhy v uzlu $T_{qi} \cong \frac{T_{si}}{1-\rho_i^m}$
 - V systému $T_q = \frac{1}{\Lambda_0} L_q$
- Pravidelnější příchody ($C_{si} < 1$)
 - „lepší než poissonovské“ → lze odhadnout nejhorší možný případ
- Složitější rozdělení, vlastnosti větvení
 - Analyzovat simulační běh



Příklad – finanční úřad



- Přichází 15 lidí za hodinu, obsluha trvá průměrně 3 minuty
 $\rightarrow \lambda = 15 \text{ hod}^{-1}, T_s = 0,05 \text{ hod}, \mu = \frac{1}{T_s} = 20 \text{ hod}^{-1}$
- Průměrně 1 formulář z 10 vyplněn špatně
 $\rightarrow p = 0,9$
- Všechny doby exponenciální (pravidelnější - lepší výsledek)

$$L_q = \frac{\rho}{(1-\rho)} \cong \frac{0,833}{0,166} \cong 5,018$$

$$L_w = L_q - \rho \cong 4,185$$

$$T_q = L_q / \lambda \cong 0,334 \text{ hod} \cong 20 \text{ min}$$



Uzavřené sítě front

- Žádné vstupy požadavků z okolí
 - Pevná množina požadavků „krouží“ sítě
→ fronty nemohou růst do nekonečna
- Hledáme typicky
 - Střední frekvenci průchodů nějakým místem (*propustnost*)
 - Střední doba průchodu mezi dvěma body v síti
- Analytické řešení jen pro poissonovské sítě
 - Lze řešit s využitím markovských modelů
 - Obecné řešení je složité

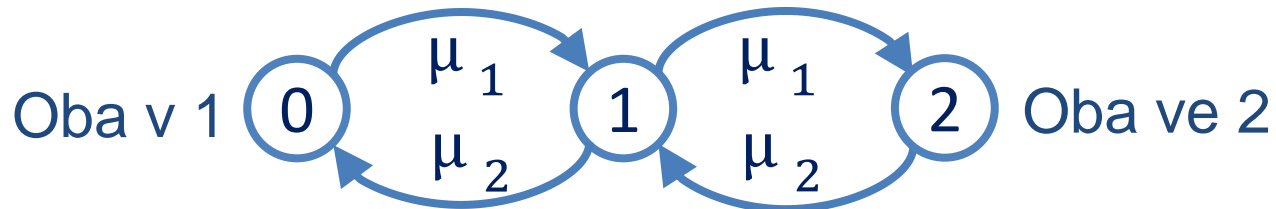
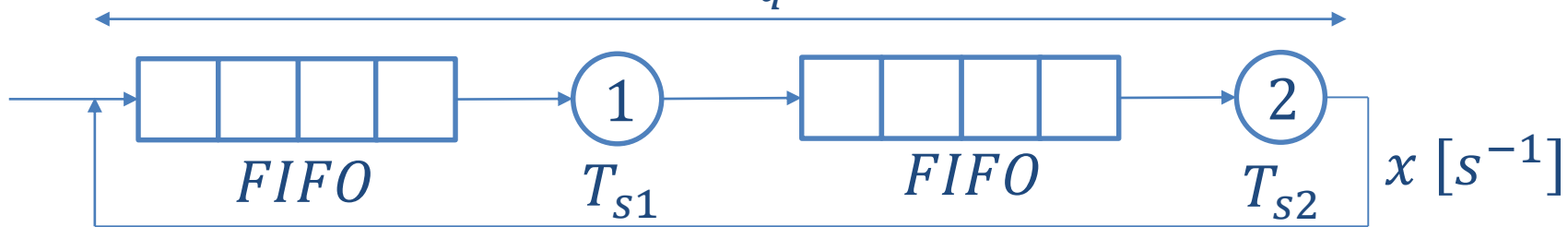


VSP - Síť front



Jednoduchý příklad

T_q

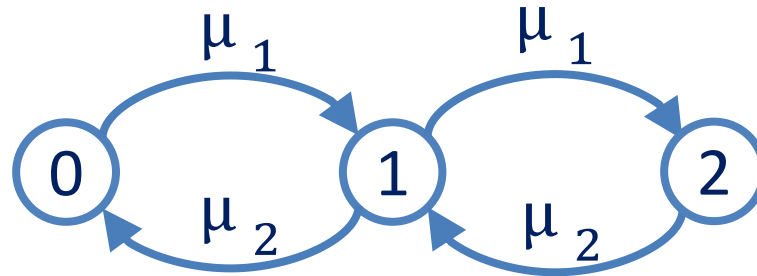


VSP - Síťe front

- 2 požadavky \rightarrow jen 3 stavy, lze určit pravděpodobnosti
- x – průchodnost
- Intenzity obsluh: $\mu_1 = 1/T_{s1}$, $\mu_2 = 1/T_{s2}$



Jednoduchý příklad



- Průchodnost

$$x = p_1 \cdot \mu_2 + p_2 \cdot \mu_2$$

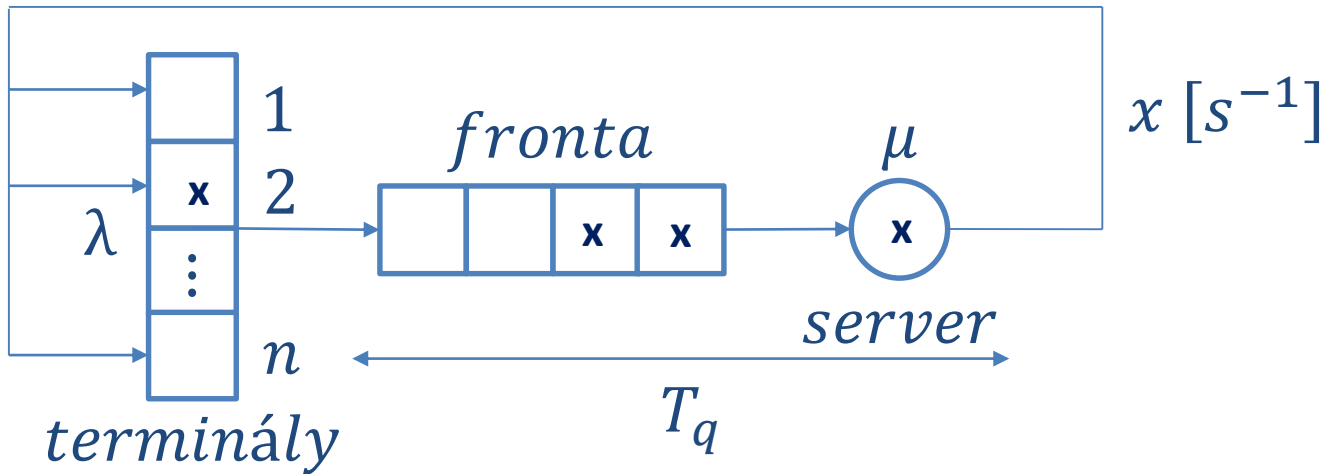
- Délky front v tabulce

$$L_{w1} = 1 \cdot p_0 + 0 \cdot p_1 + 0 \cdot p_2$$

| Stav | Fronta 1 | Fronta 2 |
|------|----------|----------|
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 2 | 0 | 1 |



Příklad – interaktivní systém

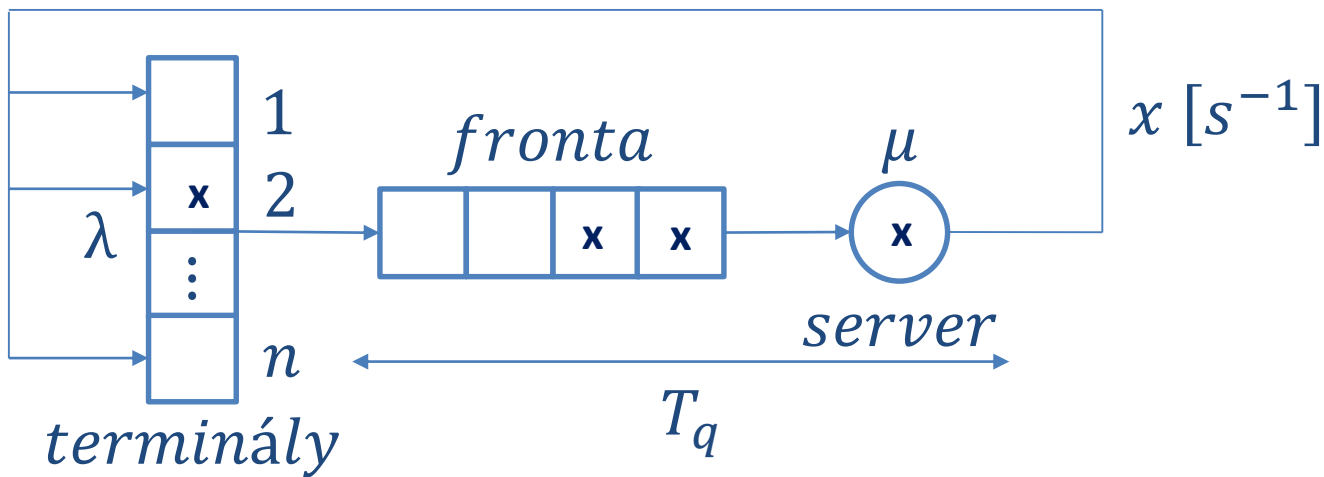


VSP - Síť front

- Server + n terminálů
 - Jednovláknové zpracování, obsluha s intenzitou μ
 - Terminály stále v provozu, uživatelé zadávají požadavky s frekvencí λ
 - Exponenciální rozdělení
- Požadavek – práce uživatele
 - Přemýšlí (λ), čeká ve frontě, čeká na výsledek (μ)



Příklad – interaktivní systém

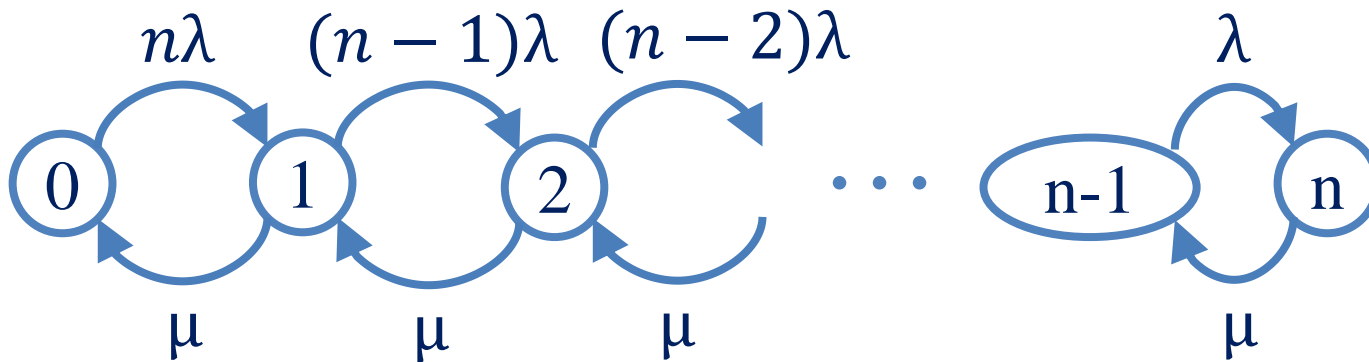
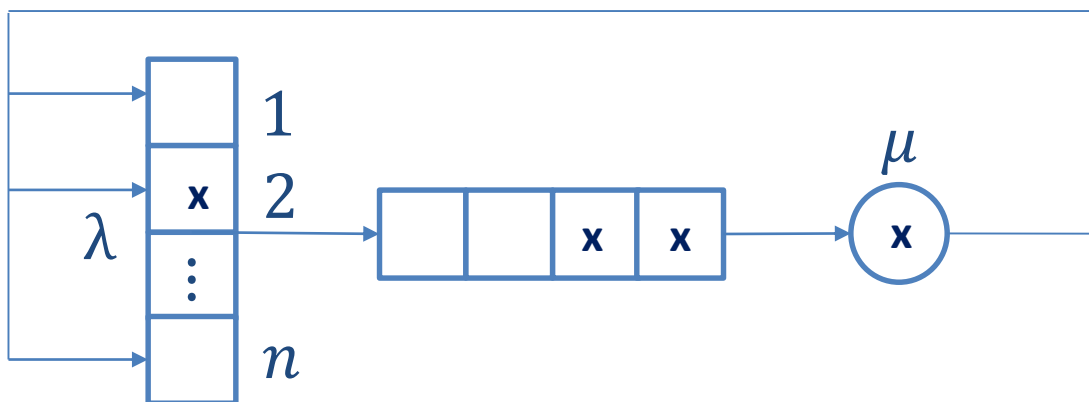


VSP - Síť front

- Střední doba přemýšlení T_t , $\lambda = 1/T_t$
- Střední doba obsluhy T_s , $\mu = 1/T_s$
- Poměr $\frac{T_s}{T_t} = \frac{\lambda}{\mu} = \rho$ – nemá význam zatížení jako v otevřených SHO
- Stav systému – počet požadavků v SHO

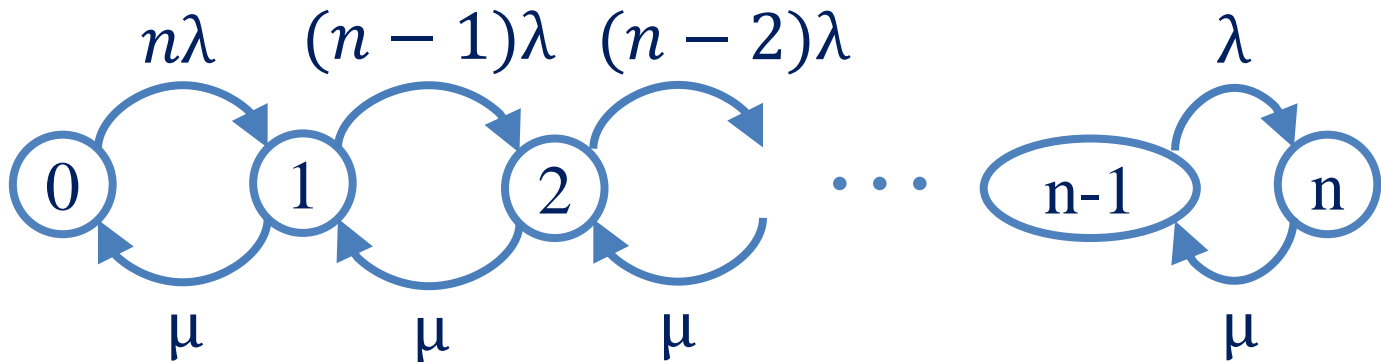


Markovský model interaktivního systému





Markovský model interaktivního systému

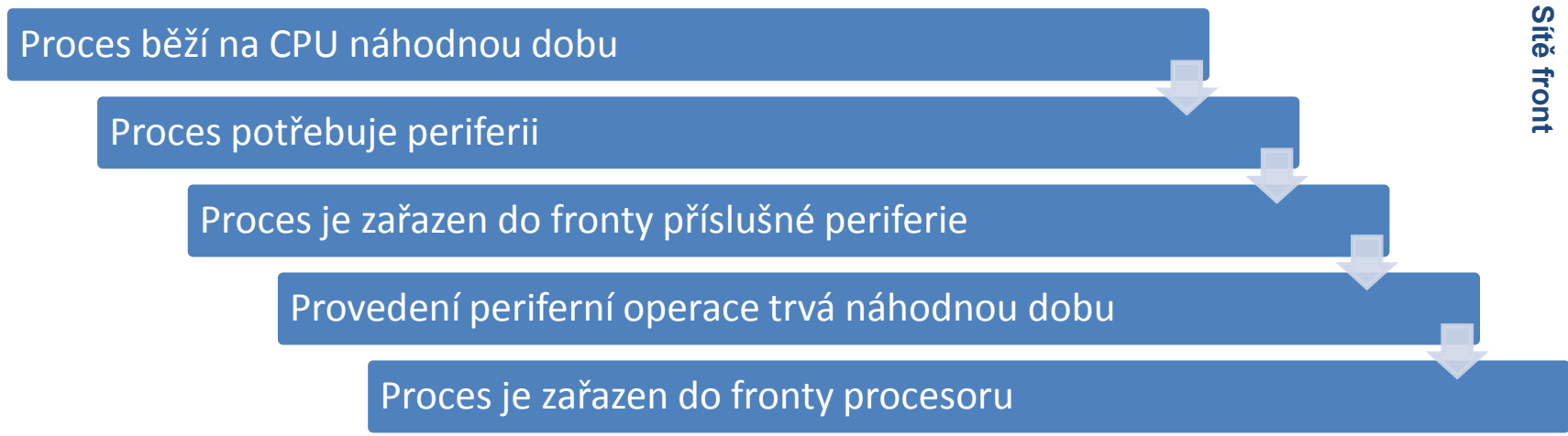


- Lze určit ustálené pravděpodobnosti (předpokládáme že existují – soustava lin. rovnic) $p_0 \dots p_n$ a z nich určit vlastnosti
- $x = \frac{1}{T_s} (1 - p_0)$ ($1 - p_0$ - pravděpodobnost že je kanál obsazený)
 - Nepodmíněná frekvence obsluhy (podmíněná - μ)
- $T_q = \left(n \frac{T_s}{1-p_0} \right) - T_t = \frac{n}{x} - T_t$



Příklad – OS s více procesy

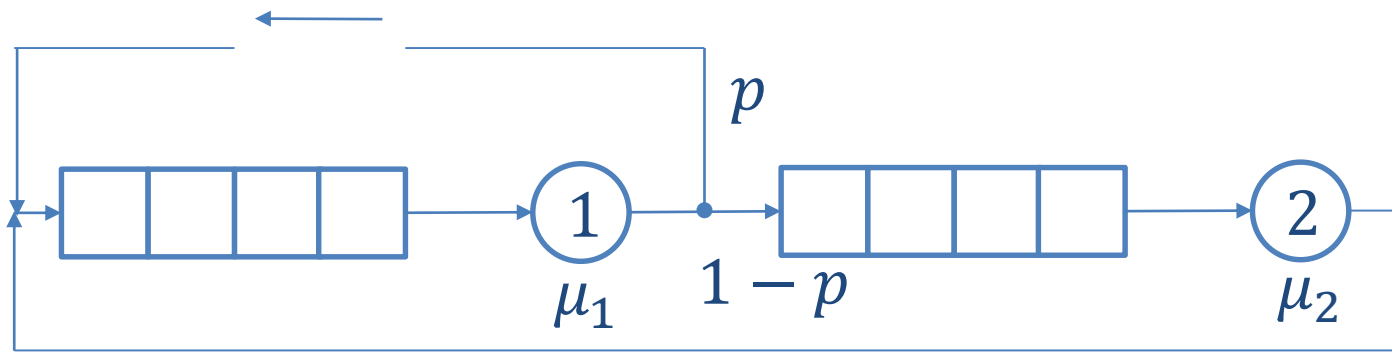
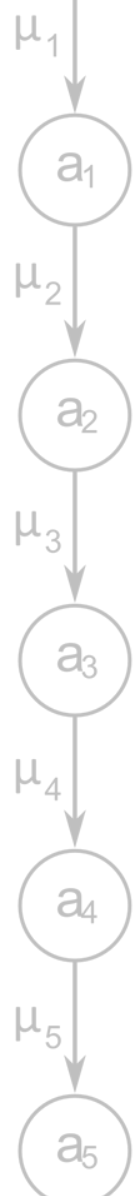
- 2 elementární SHO:
 - Procesor – elementární SHO
 - Řadič periferií
- Pevně daný počet najednou zpracovávaných úloh
- Ukončený proces je ihned nahrazen dalším (dávka požadavků ve vnější paměti)
 - počet průchodů přes procesor = propustnost





Konkrétní příklad – správa paměti

- Součásti:
 - CPU (1)
 - Řadič paměti (stránkování, 2)
- 3 procesy najednou, stejné vlastnosti
- p – pravděpodobnost že opuštění procesoru znamená ukončení programu
- Exponenciální rozdělení
 - Dobré pro konflikty stránek, pro disk lépe Gaussovské nebo rovnoměrné

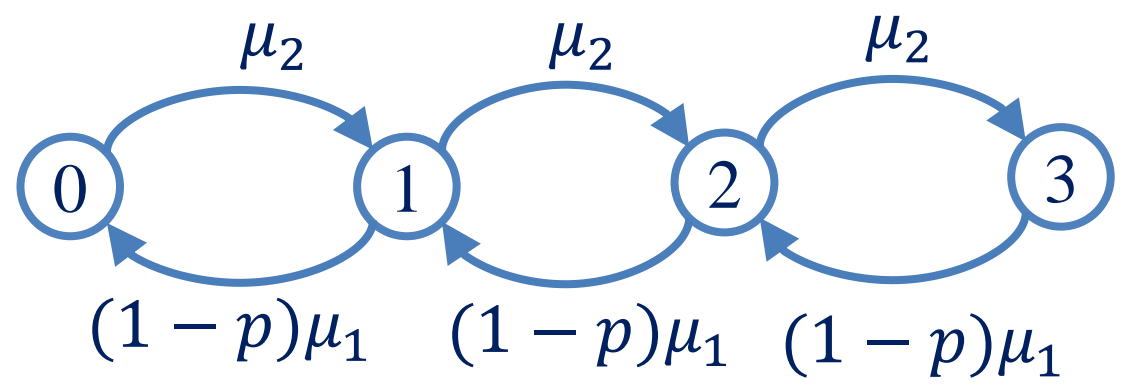
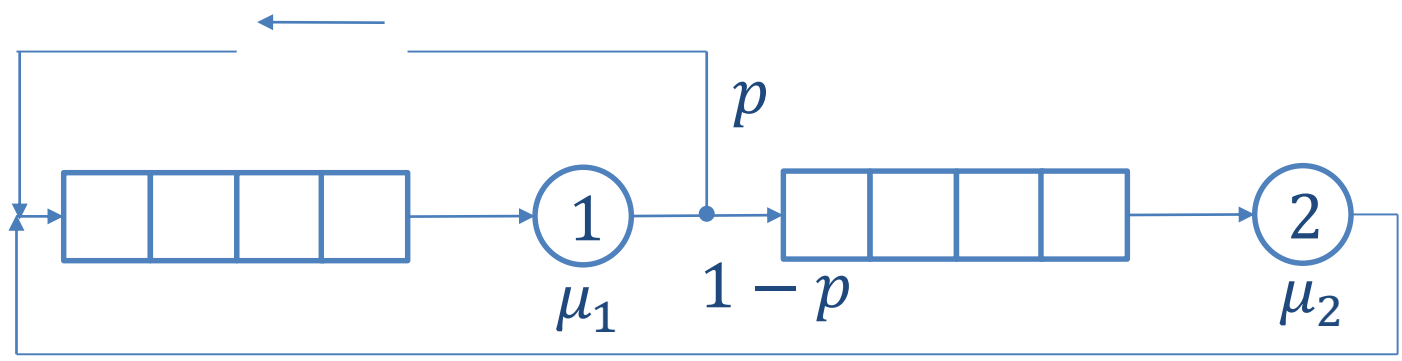


VSP - Síť front



Konkrétní příklad – správa paměti

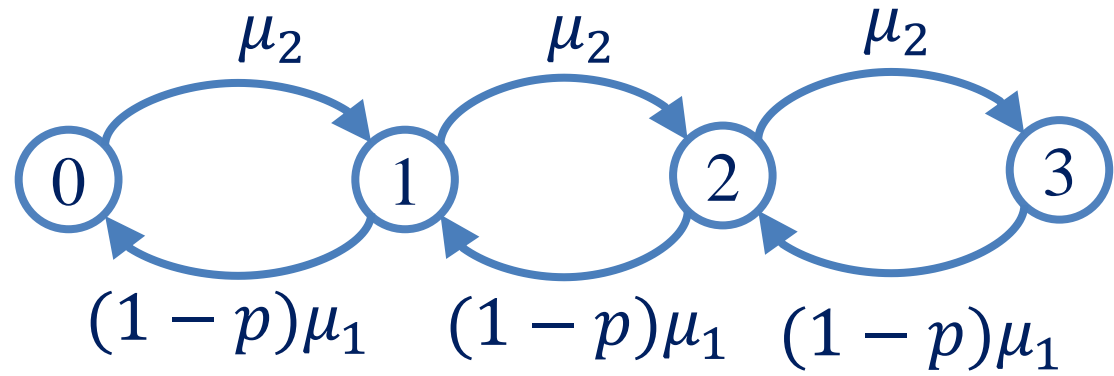
- Lze sestavit markovský model bez absorbčních stavů



VSP - Síť front



Konkrétní příklad – správa paměti

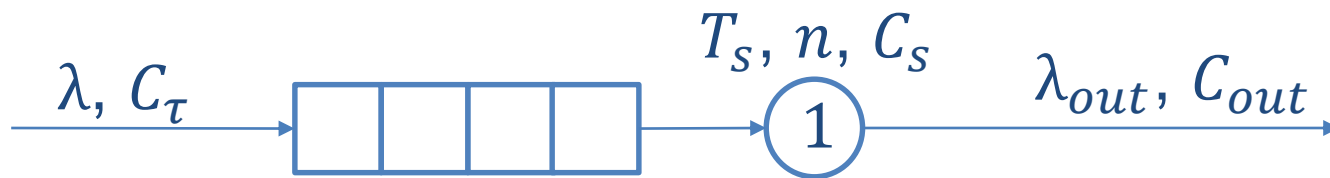


- Průchodnost $x = \mu_1 p (p_1 + p_2 + p_3)$
 - p – pravděpodobnost ukončení procesu
 - Může nastat jen když je aktuálně nějaký proces v CPU
 - Neodpovídá žádné explicitní hraně v markovském modelu
- Střední doba obsluhy $T_q = \frac{n}{x} = \frac{3}{x}$
 - Proces není nikde zdržen, neodečítáme žádné T_t



Nepoissonovské sítě front – přibližné řešení

- Obecné toky \rightarrow charakterizovány frekvencí λ a koeficientem variace C_τ
- Otevřená síť front



- $\lambda_{out} = \lambda$ (jinak nebude stacionární)
- $C_{out}^2 = 1 + \rho^2(C_s^2 - 1) + (1 - \rho^2)(C_\tau^2 - 1)$
– ρ určuje vliv rozdělení dob vstupů a dob obsluh
- Nelze snadno sčítat a dělit toky



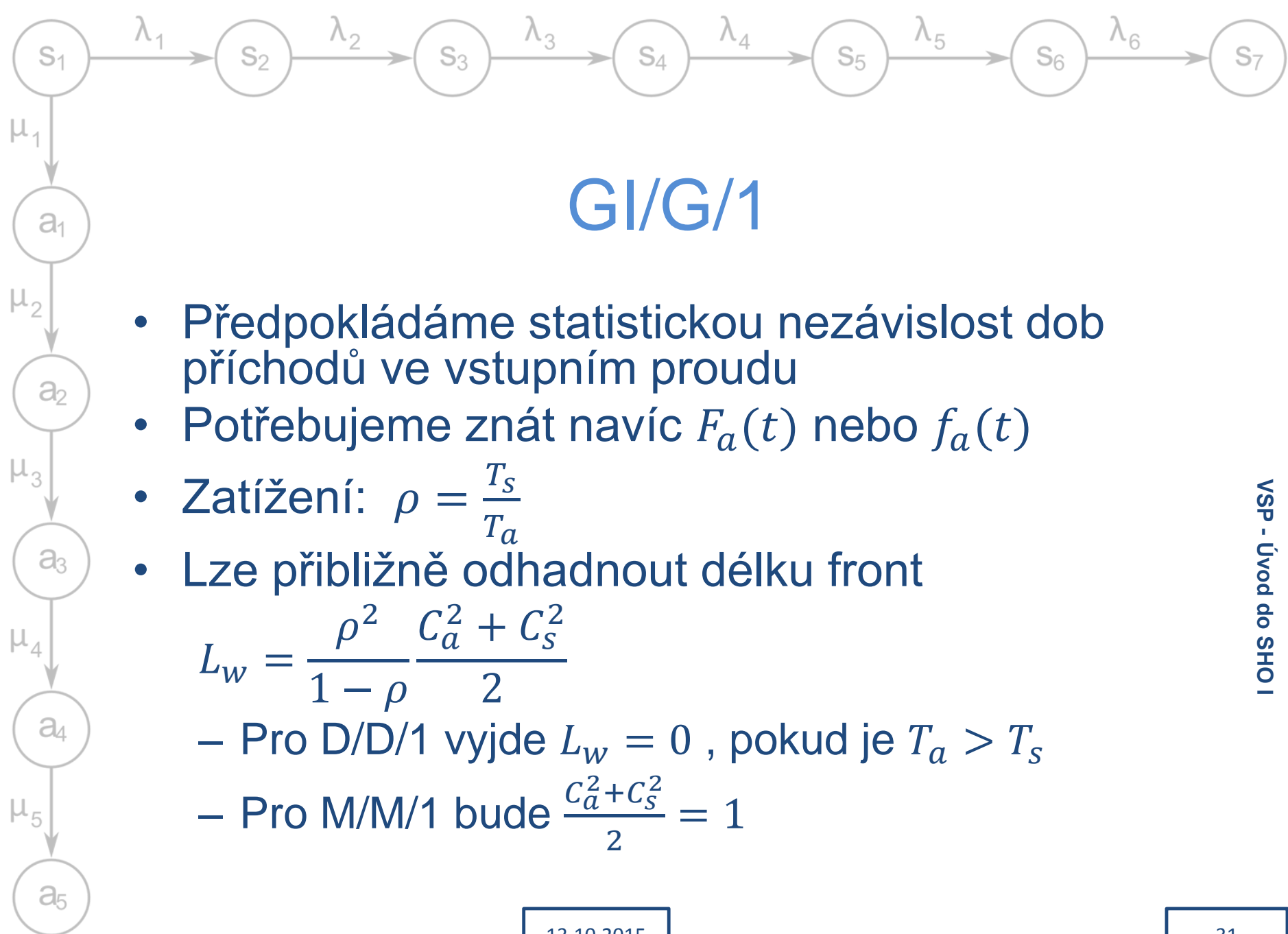
VSP - Síť front



Nepoissonovské sítě front – toky

- Dělení toku s λ, C_τ na n toků
 - $\lambda_i = \lambda p_i$
 - $C_i^2 = 1 + p_i(C_\tau^2 - 1)$
- Slučování toků s λ_i, C_i
 - $\lambda = \sum_{i=1}^n \lambda_i$
 - $C^2 = 1 + \sum_{i=1}^n \left(\frac{\lambda_i}{\lambda}\right)^2 (C_i^2 - 1)$
- Pro jednotlivé elementární SHO lze použít vztahy pro GI/G/n

VSP - Sítě front





Samostatná práce

- Síť s alespoň 4 uzly a 2 zpětnými vazbami
 - Nerozložitelná – zpětné vazby přes celý systém
 - Návrh na přednášku 10. listopadu
- Parametry volit tak, aby byla ve stacionárním režimu, ale se zátěží ($\rho > 0,5$)
- Určit toky v uzlech, L a T parametry v uzlech i pro celou síť
- Porovnat vypočtené a odsimulované hodnoty
- Sledovat další charakteristiky
 - Počet požadavků které projdou zadaným uzlem
 - Charakteristiky toku mezi vybranými uzly (histogram, střední hodnota, směrodatná odchylka)
 - Průměrný počet obrátek v systému
 - ...



Děkuji za pozornost



- Příště simulační analýzy SHO