454-304/1: Spojovací soustavy

Základy teorie hromadné obsluhy

Miroslav Vozňák
VŠB - Technical University of Ostrava
Department of Telecommunications
Faculty of Electrical Engineering and Computer Science
17. listopadu 15, 708 33 Ostrava – Poruba

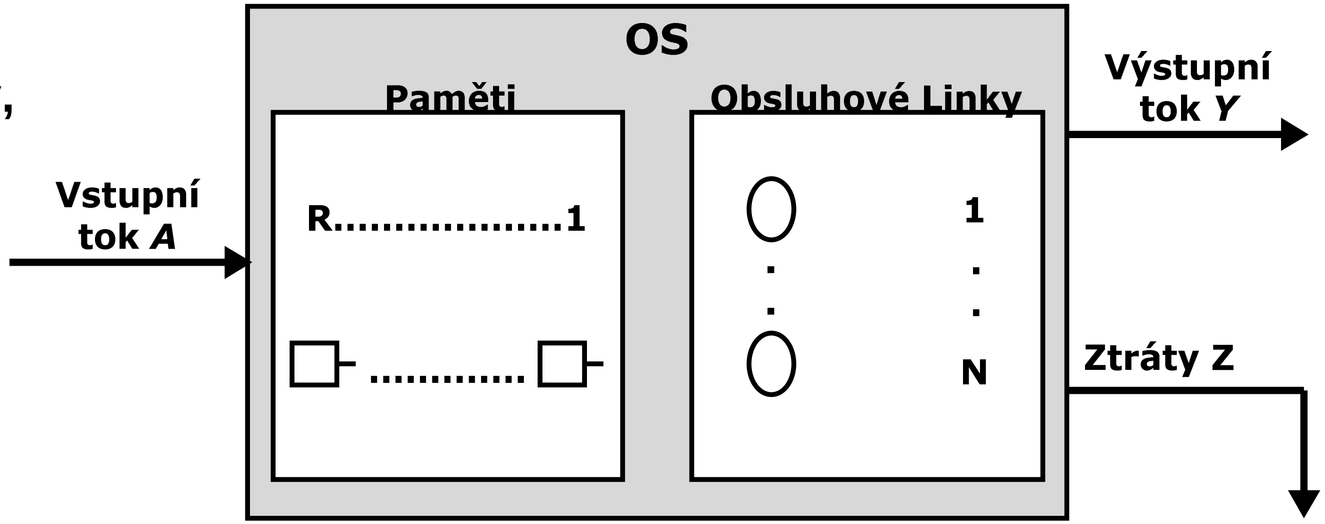
mailto:miroslav.voznak@vsb.cz
http://homel.vsb.cz/~voz29

Obsluhový systém a parametry

Obsluhový systém (OS) slouží pro uspokojení vznikajících požadavků v dohodnuté kvalitě. Kvalitu OS vyjadřuje obecně míra jeho pohotovosti vyhovět žádosti v plném, nebo částečném rozsahu (pokud se týče doby obsluhy).

Žádosti vytváří **provozní toky**
-vstupní tok o intenzitě A
-tvořený požadavky z *s* zdrojů
-výstupní tok intenzity Y.
-neuspokojené žádosti tvoří ztráty s hustotou Z.

Vstupní tok
-**deterministický,**
-**stochastický**



Paměť je místem v OS, kde je umožněno čekání žádostem. V **rozdělení obsluhových systémů** na:

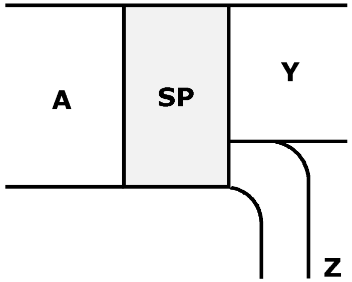
- pracující se ztrátami,
- s čekáním a ztrátami,
- s čekáním.

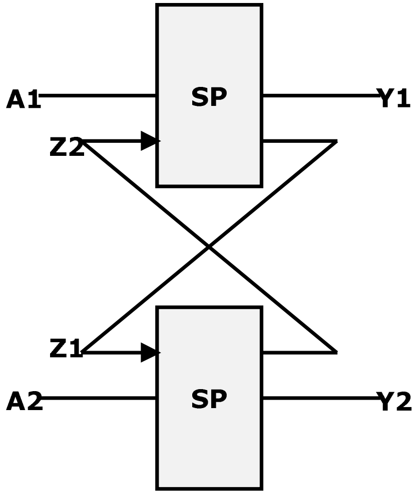
Provozní zatížení

- základní kvantitativní parameter všech OS.
- obsazení všech linek v příslušném místě OS za dobu T, časovou jednotkou byla stanovena 1 hod.
- celková doba obsazení N obsluhových linek, nechť Nx je průběh jejich obsazení v čase, Nx vyjadřuje počet použitých spojovacích cest v okamžiku t, potom platí:

$$Y = \frac{1}{T} \int_{t_0}^{t_0+T} N_x \cdot dt \qquad [erl]$$

Pro intenzitu provozního zatížení byl stanoven jednotkou *erlang* [erl]. Jedna obsluhová linka může zpracovat intenzitu max. 1 erl, pokud je trvale obsazena po dobu 1hod

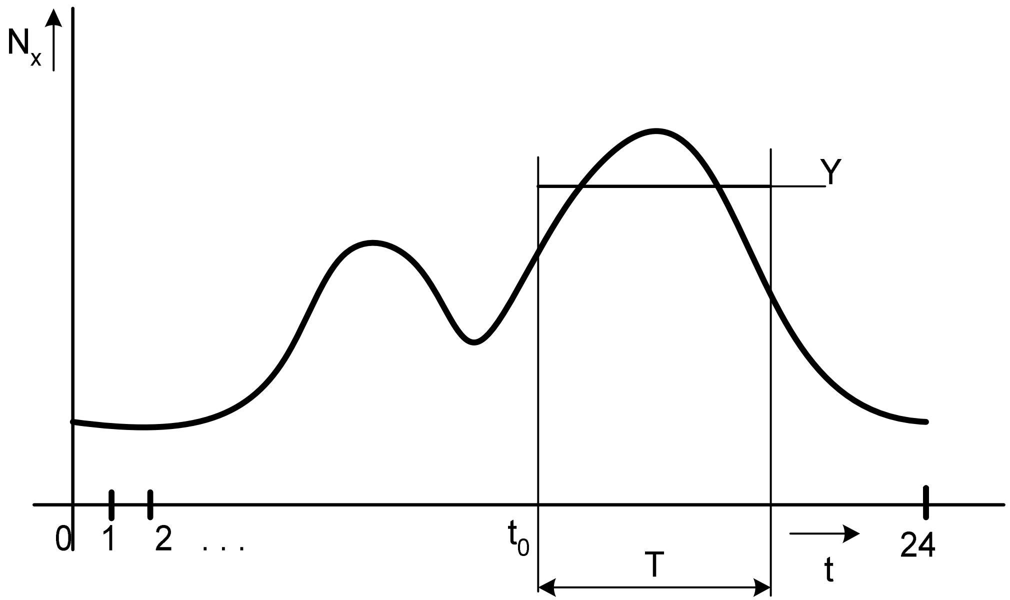




Na vstupu OS označme intenzitu nabízeného provozního zatížení A, na výstupu OS bude intenzita přeneseného provozního zatížení Y. Odstupující provozní zatížení vyjádříme intenzitou Z, potom můžeme tvrdit: $Y = A – Z$ [erl]

přeliv mezi svazky – ztracené zatížení jednoho svazku je nabídnuto na vstup jiné svazku

Definujme hlavní provozní hodinu **HPH** jako čtyři po sobě jdoucí 15-ti minutové intervaly s největším provozním zatížením, Měření probíhá po čtvrthodině a vyhodnocuje se nejvyšší dosažený součet čtyř po sobě jdoucích intervalů.



BHT (Busy Hour Traffic) je hodina nejvyššího provozu (HPH),
BHCA (Busy Hour Call Attempts) je hodina s největším počtem volání,
BHCC (Busy Hour Call Completions) je hodina s největším počtem sestavených vol.,
CPS (Call per Second) je počet volání za sekundu,
AHT (Average Hold Time) je průměrná doba čekání.

V HPH je zpracováno cca 15% celkového denního zatížení. Průměrná doba hovoru se pohybuje kolem dvou minut, přitom více než polovina hovorů je kratších než 30 vteřin. Označme provozní zatížení přenesené za den jako Y_D, potom poměr přeneseného zatížení v HPH k celkovému dennímu se označuje jako koncentrace k.

$$k = \frac{Y_{HPH}}{Y_D}$$

Nabízené zatížení můžeme získat následovně:

s je počet účastníků,
c_0 je počet volání na účastníka za den,
t_0 je střední doba obsazení spojovací cesty při jednom volání
k je koncentrace.

$$A = s \cdot c_0 \cdot t_0 \cdot k$$

Pro OS, který neumožňuje vytváření front, tj. počet míst ve frontě R=0, definujeme ztráty Z jako podíl

$$Z = \frac{Cz}{Cn}$$

c_z je počet ztracených požadavků, v procentech, %.
c_n je počet všech požadavků, v procentech, %.

Erlang B

Tento OS nazýváme OS se ztrátami a je prezentován telefonními ústřednami. Základní vztah mezi parametry OS se ztrátami udává první Erlangova rovnice E_1

$$E_1 = p(z) = z = \frac{\frac{A^N}{N!}}{\sum_{i=0}^{N} \frac{A^i}{i!}}$$

Erlang C

Druhou základní skupinu OS tvoří OS s možností čekání v konečné frontě, nebo v neomezeném počtu míst pro čekání. Průměrná doba čekání je hlavním kvalitativním parametrem systémů s čekáním.

Pro OS s neomezenou frontou, trpělivými žádostmi a požadavky na vstupní tok shodnými pro OS se ztrátami platí druhá erlangova rovnice E_2, vyjadřující pravděpodobnost čekání $p_č$.

$$E_2 = p(c) = \frac{\frac{A^N}{N!}\,\frac{N}{N-A}}{\sum_{i=0}^{N-1}\frac{A^i}{i!} + \frac{A^N}{N!}\,\frac{N}{N-A}}$$

Střední doba čekání žádostí $t_č$ je

$$t_c = p(c) = \frac{t_{os} \cdot E_2}{N-A}$$

t_{os} je střední doba obsluhy.

Engsetův model

Klasifikace nejpoužívanějších OS je následující:

Erlang B se užívá, jestliže počet zdrojů zatížení >> počet vedení, výsledkem je pravděpodobnost blokování, použití ve veřejných sítích,

Erlang C stejně jako v předchozím předpokládá, že počet zdrojů zatížení >> počet vedení, výsledkem je pravděpodobnost čekání ve frontě, použití pro call centra,

Engsetův model předpokládá konečný počet zdrojů, výsledkem je pravděpodobnost blokování, používá se pro PBX.

Označme *s* jako počet zdrojů, kde každý z nich vytváří tok volání o intenzitě . Pro počet obsluhových linek N můžeme vyjádřit pravděpodobnost ztráty jako .

$$E_N(\alpha) = \frac{\binom{s-1}{N}\alpha^N}{\sum_{i=0}^{N}\binom{s-1}{i}\alpha^i}$$

Za podmínky nekonečného počtu zdrojů se dá z Engsetovy rovnice odvodit ERLANG B a má tedy obecnější platnost.

Kendallova klasifikace SHO

V roce 1953 David George Kendall publikoval značení obsluhových systémů ve formě písmen **A / B / C** , kde

A (Arrival process) představuje distribuční funkci popisující intervaly mezi příchody,

B (Service Time distribution) je distribuční funkce doby obsluhy,

C (Number of Channels) je počet obsluhových linek.

Na pozicích A a B se používají písmena, která charakterizují typ distribuční funkce, těmi jsou:

M (Markovian) značí exponenciální rozložení intervalů, pro distribuční funkci F, hustotu pravděpodobnosti f, střední hodnotu E a rozptyl D platí:

$$F(t) = 1 - e^{-\lambda t} \qquad t \geq 0, \quad 0 < \lambda < \infty$$

$$= 0 \qquad t < 0,$$

$$f(t) = \lambda \cdot e^{-\lambda t} \qquad , 0 < t < \infty$$

$$E[t] = \frac{1}{\lambda}$$

$$D[t] = \frac{1}{\lambda^2}$$

D (Degenerate) představuje deterministicky stanovené rozložení intervalů,

$$F(t) = 1 \qquad t \geq \frac{1}{\lambda}, \quad 0 < \lambda < \infty$$
$$= 0 \qquad t < \frac{1}{\lambda},$$
$$E[t] = \frac{1}{\lambda}$$
$$D[t] = 0$$

Ek (Erlang distribution) označuje Erlangovo rozložení k-tého řádu,

$$F_k(t) = 1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t} \qquad t \geq 0, \quad 0 < \lambda < \infty$$
$$= 0 \qquad t < 0,$$
$$f_k(t) = \frac{(\lambda \cdot t)^{k-1}}{(k-1)!} \lambda \cdot e^{-\lambda t} \qquad , 0 < t < \infty$$
$$E[t] = \frac{k}{\lambda}$$
$$D[t] = \frac{k}{\lambda^2}$$

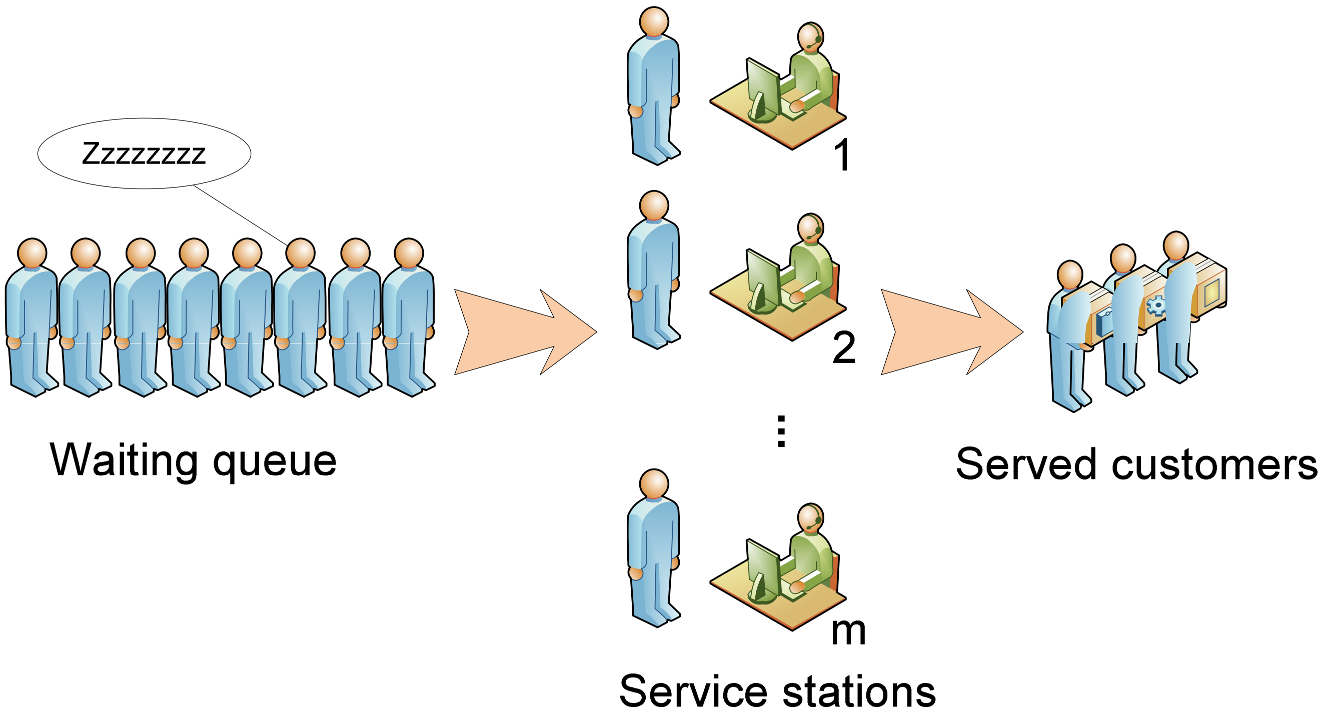
Později bylo Kendallovo značení doplněno o další charakteristiky obsluhového systému na konečný tvar:

$$A/B/C/K/N/D$$

K (number of places in the system) je maximální počet požadavků v systému představujících počet míst pro čekající anebo počet obsluhových linek plus počet míst pro čekající (bohužel je zde nejednoznačnost),
N (calling population) je počet zdrojů požadavků, neuvádí se pokud je nekonečný,
D (queue’s discipline) je režim fronty jako FCFS (First Come - First Served), FIFO (First In - First Out), FCLS (First Come - Last Served), LIFO (Last In - Last Out), SIRO (Service In Random Order).

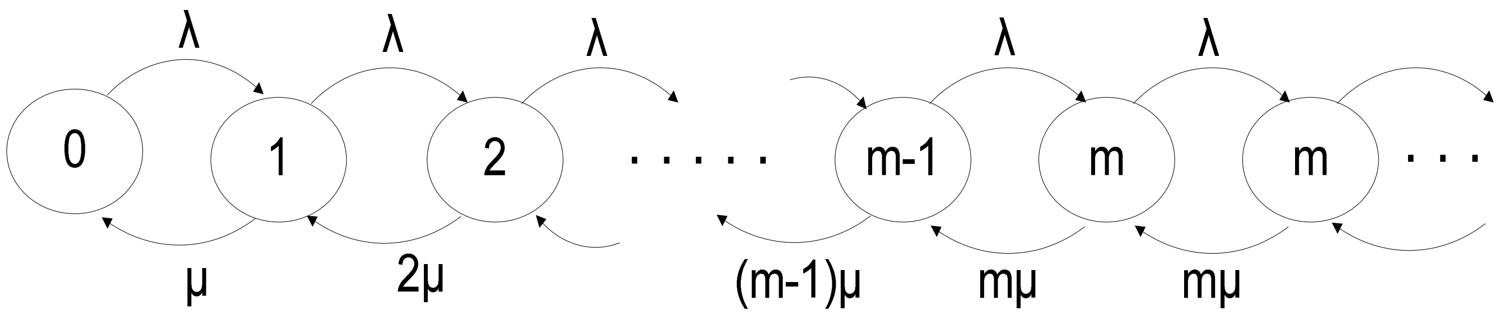
Původní Kendallovo značení $A/B/C$ předpokládá, že $K = \infty$, $N = \infty$ a D = FIFO

Markov Model M/M/m/∞ in Contact Center Environment



- incomming calls: Poisson distribution,
- service time: Exponential distribution.

Markov Model M/M/m/∞



Important traffic parameters

System stability

$$\rho = \frac{\lambda}{m\mu}$$

Probability of waitinq in waiting queue

$$P_Q = \frac{(m\rho)^m}{m!(1-\rho)} P_0$$

Probability of empty system

$$P_0 = \left\{ \sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!} \left(\frac{1}{1-\rho} \right) \right\}^{-1}$$

Important traffic parameters

Number of requests in the system

$$N = \rho m + \frac{\rho}{(1-\rho)} P_Q$$

Mean number of requests waiting in the waiting queue

$$Q = \frac{\rho}{(1-\rho)} P_Q$$

Mean time spend in the waiting queue

$$W = \frac{\rho}{(1-\rho)} \frac{P_Q}{\lambda}$$

Grade of Service

- ratio of incomming calls that will be assigned to agent until time T_W

$$GoS = 1 - P_Q e^{-\mu(m-A)T_W}$$

Calculation of important traffic parameters

Let assume:

- λ = 60 (incomming calls per hour)
- T_{serv} = 5 minutes (call handling time)
- A = 5 Erl (average traffic load)
- P_Q = 5 %

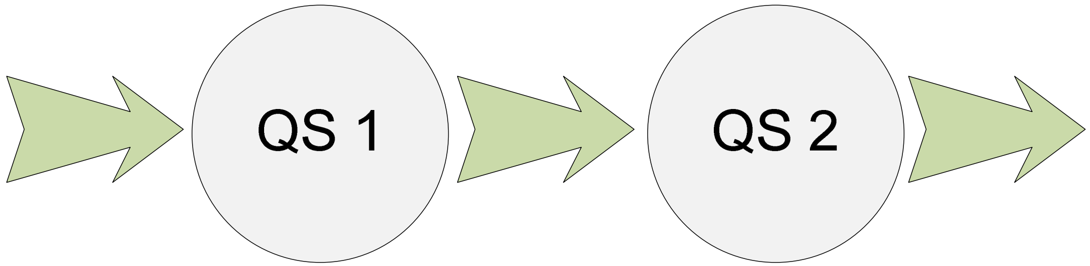
Calculation of required number of agents

m	**P_Q** [%]	**N**	**T** [min]	**Q**	**W** [min]	**GoS** [%]	**ρ**
8	16,73	5,28	5,28	0,28	0,28	87,61	0,63
9	8,05	5,10	5,10	0,10	0,10	94,60	0,56
10	3,61	5,04	5,04	0,04	0,04	97,81	0,50
11	1,51	5,01	5,01	0,01	0,01	99,17	0,45
12	0,59	5,00	5,00	0,00	0,00	99,71	0,42

traffic parameters if A=5 Erl

Contact center as multiple-phase queuing system

- Contact center in which the first contact with customer is made by IVR

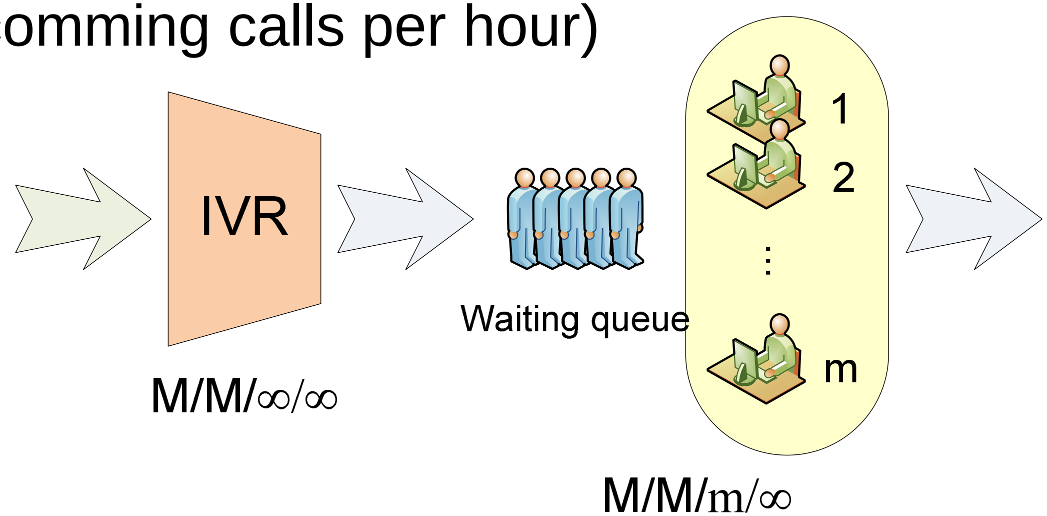


- QS1: IVR modeled by Markov model M/M/∞/∞
- QS2: modeled by Markov model M/M/m/∞

Contact center as multiple-phase queuing system

Let assume:

- T_{serv1} = 1 min. (request in IVR)
- T_{serv2} = 4 min. (call handling time by agent)
- T_{serv} = 5 minutes (call handling time)
- λ = 60 (incomming calls per hour)



Impact of IVR on modeled Contact Center

	A [Erl]	N_1	N_2	N	m	P_Q [%]	Savings [%]
without IVR	5	–	–	5,04	10	3,61	–
with IVR	5	1,00	4,06	5,06	8	5,90	20
with IVR	5	1,00	4,02	5,02	9	2,38	10

comparison of systems with/without IVR

Skripta Spojovací systémy: ***Kapitola*** *6*

Děkuji za pozornost

miroslav.voznak@vsb.cz