



# Badania operacyjne

## Systemy masowej obsługi

### 1. Opis i działanie

Jerzy Konorski  
pok. 139 EA  
jekon@eti.pg.gda.pl

zaliczenie tej części: test końcowy (min 17 pkt z 34)

*J. Konorski, Badania operacyjne/Systemy masowej obsługi / I*

1



### Literatura

- L. Kleinrock: *Queuing systems*, vol. I, II, Wiley 1975-1976
- D. Gross, C.M. Harris: *Fundamentals of Queuing Theory*, Wiley 1998
- Joti Lal Jain, W. Boehm, Sri Gopal Mohanty: *A Course on Queuing Models*, Chapman & Hall 2006
- G. Bolch, S. Greiner, H. de Meer, K. S. Trivedi: *Queueing Networks and Markov Chains. Modeling and Performance Evaluation with Computer Science Applications*, 2nd Ed., Wiley-Interscience 2006
- D. Koenig, D. Stoyan: *Metody teorii obsługi masowej*, WNT, Warszawa 1979
- G.P. Klimow: *Procesy obsługi masowej*, WNT, Warszawa 1979
- J. Blazewicz i in.: *Badania operacyjne dla informatyków*, WNT, Warszawa 1983, rozdziały 3 i 4
- B. Filipowicz: *Modele stochastyczne w badaniach operacyjnych. Analiza i synteza systemów obsługi i sieci kolejkowych*, WNT, Warszawa 1996
- T. Czachórski: *Modele kolejkowe w ocenie efektywności sieci i systemów komputerowych*, Wyd. J. Skalmierski, Gliwice 1999

2

*J. Konorski, Badania operacyjne/Systemy masowej obsługi / I*



## Systemy masowej obsługi

- **system informatyczny** (komputer / call center / baza danych / serwer WWW) : przerwania / uaktualnienia / zapytania / transakcje oczekują w kolejce na zwolnienie się operatorów / jednostek obliczeniowych / obszarów danych
- **urządzenie telekomunikacyjne** (karta sieciowa / centrala telefoniczna / koncentrator szybkiego łącza danych):
- **infrastruktura transportowa** (wjazd na autostradę / stacja benzynowa / nabrzeże portowe / pas startowy): pojazdy oczekują na "szczelinę obsługi" (slot)
- **punkt świadczenia usług** (bankomat / kasa w supermarketie / urząd): klienci / petenci oczekują na zwolnienie się osób / urządzeń obsługujących

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

3

### Systemy masowej obsługi (2)

*przeływy strumień ruchu router*

*POST OFFICE*

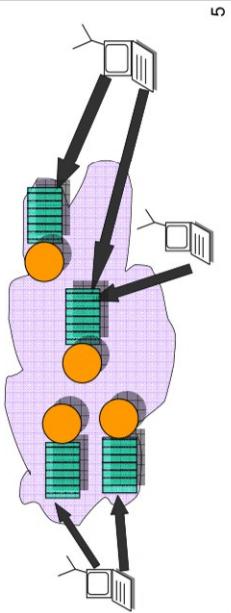
Wszędzie kolejki!  
Trzeba je uwzględniać w projektowaniu systemów.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

4

## Systemy masowej obsługi (3)

- system posiada **zasoby** – organiczne, zwrotne
  - reaguje na zdarzenia w postaci przybywających **zgłoszeń** = żądań dostępu do zasobów
  - w odpowiedzi przydziela zgłoszeniu wybrane zasoby, umożliwiając ich konsumpcję przez pewien skończony czas = **obsługa zgłoszenia**
  - zasoby systemu zdolne do świadczenia obsługi = **procesory**
  - zgłoszenie może natrafić na procesory zajęte obsługą innych zgłoszeń
  - w takim wypadku zostaje umieszczone w **pamięci buforowej**, gdzie oczekuje w kolejce na zwolnienie się procesora i rozpoczęcie obsługi



J. Komorski, Badania operacyjne/Systemy masowej obsługi I

5

## Misja teorii kolejek

Populacja zgłoszeń / źródło zgłoszeń jest zwykle bardzo liczna.

Pozbawia to sensu opymalizację konkretnych scenariuszy przybywania zgłoszeń, np. harmonogramowanie zbioru zadań minimalizujące czas zakończenia lub liczbę wykorzystanych procesorów.

Można jedynie analizować i projektować zasady obsługi przybywających "na bieżąco" strumieni **zgłoszeń**. W tym celu studujemy przebieg czasowe rozmaitych charakterystyk kolejek = **procesy kolejkowe (procesy obsługi)**

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

6



## Misja teorii kolejek (2)

Przy licznej populacji zgłoszeń chwilowy popyt na obsługę często przekracza chwilową podaż obsługi – wtedy tworzą się kolejki.  
Należy ograniczać ich negatywne skutki, np.

- niezadowolenie klientów (opóźnienia, odmowa / rezygnacja z obsługi),
- koszty buforowania i zarządzania,

dbając jednocześnie o ekonomiczność wykorzystania procesorów.

Narzędzia badawcze po temu wypracowane zostały przez ważny dział badań operacyjnych – **teorię kolejek** (inaczej **teorię masowej obsługi**).

Zaczęło się podczas II wojny światowej, od bombowców krających w kolejce do lądowania...

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

7



## Najprostszy model

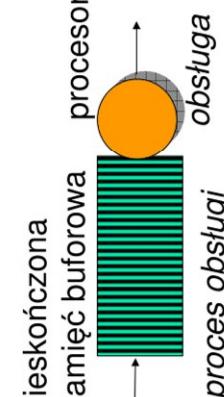
strumień zgłoszeń



generacja

Elementy najprostszego modelu systemu masowej obsługi:

- pamięć buforowa
- procesor
- strumień zgłoszeń



Charakterystyki procesu obsługi zależą od charakterystyk strumienia zgłoszeń i zasad pracy systemu pamięć buforowa-processor.

Jak?

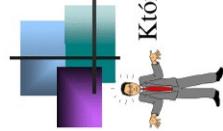
Tym właśnie zajmuje się teoria kolejek.

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

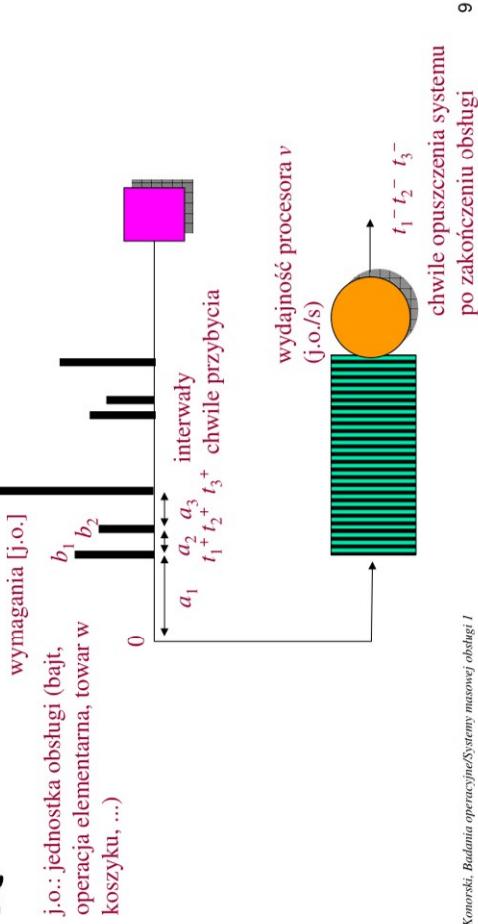
8



## Najprostszy model (2)



Które charakterystyki strumienia zgłoszeń i zasady obsługi są istotne?



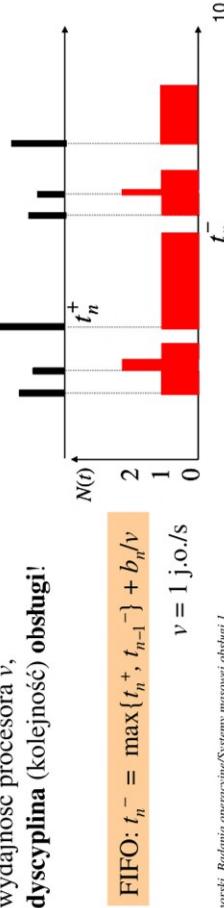
## Najprostszy model (3)



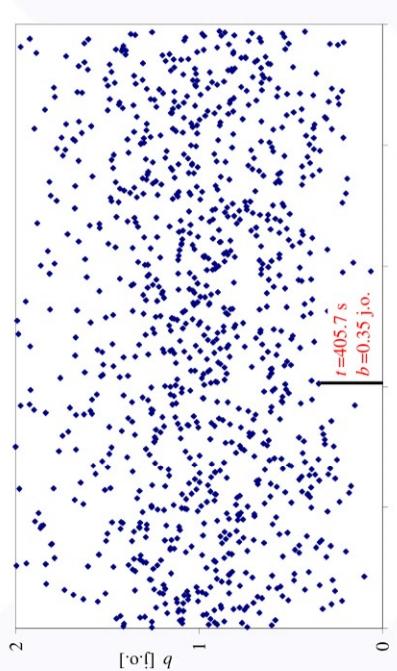
Uwagi:

- $I_n^+ = a_1 + \dots + a_n$
- $b_n/v = \tau_n =$  wymagany czas obsługi zgłoszenia  $n$
- $d_n = t_n^- - t_n^+ =$  opóźnienie systemowe zgłoszenia  $n$
- $w_n = d_n - b_n/v =$  opóźnienie buforowania zgłoszenia  $n$  (stracony czas)
- $N(t) = \#\{n \mid t_n^+ \leq t \leq t_n^-\} =$  liczba zgłoszeń w systemie w chwili  $t$

Dla konstrukcji procesów  $N(t)$  i  $w_n$  istotne są:  
 - wydajność procesora  $v$ ,  
 - dyscyplina (kolejność) obsługi!



## Strumień zgłoszeń i proces obsługi – przykład



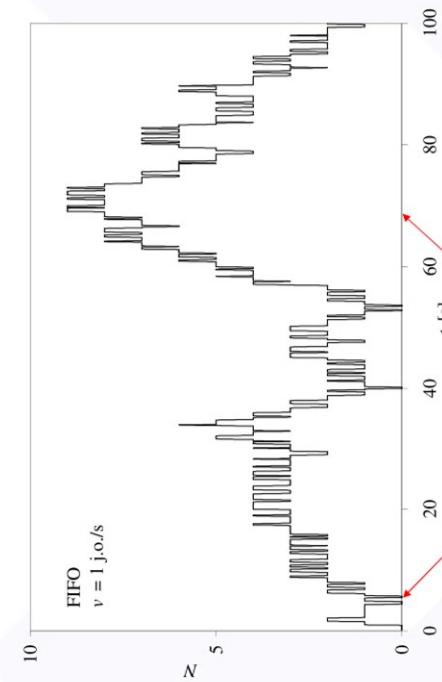
średni interwał  $a_{\text{sr}} = 1$  s

średnie wymaganie  $b_{\text{sr}} = 1$  j.o.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

11

## Strumień zgłoszeń i proces obsługi – przykład (2)

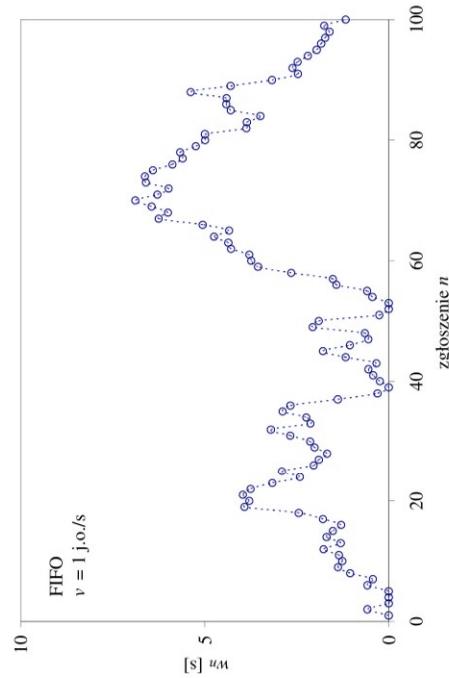


okresy bezczynności i zajętości procesora

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

12

### Strumień zgłoszeń i proces obsługi – przykład (3)



J. Komorski, Badania operacyjne/Systemy masowej obsługi I

13

### Jakie właściwości ma "dobry" proces obsługi?

- z punktu widzenia zgłoszeń: niewielkie opóźnienia buforowania, rzadkie odmowy przyjęcia zgłoszenia wskutek zbyt długiej kolejki
- z punktu widzenia operatora systemu: wysokie wykorzystanie procesora (rzadkie okresy bezczynności)

Sprzeczne oczekiwania! Coraz rzadszym okresem bezczynności towarzyszy:

- występowanie kolejek
- wzrost długości kolejek
- systematyczne narastanie kolejki / lawina odmów przyjęcia  
– procesor "nie wyrabia się z obsługi" w czasie rzeczywistym!

Relacja pomiędzy charakterystykami strumienia zgłoszeń a wydajnością procesora określa **obciążenie systemu**.  
Jak je ilościowo zdefiniować?  
Jaką to daje informację o działaniu systemu?  
Przy jakim obciążeniu procesor przestaje "wyrabiać się z obsługi"?

14



J. Komorski, Badania operacyjne/Systemy masowej obsługi I

## Współczynnik obciążenia – ruch oferowany

Rozpatrzmy d盲ugi okres obserwacji  $T$ .

W tym czasie przybędzie w przybliżeniu  $T/a_{\text{sr}}$  zgłoszeń, łącznie reprezentują one średni **popyt na obsługę** =  $b_{\text{sr}}(T/a_{\text{sr}})$  j.o.

**Podaż obsługi** w tym czasie wyniesie  $vT$ .

**Współczynnik obciążenia** systemu = stosunek średniego popytu do podaży:

$$r = \frac{b_{\text{sr}}(T/a_{\text{sr}})}{vT} = \frac{b_{\text{sr}}/v}{a_{\text{sr}}} = \frac{b_{\text{sr}}/a_{\text{sr}}}{v} \quad (\text{bezwymiarowy})$$

= średni czas obsługi zgłoszenia ( $b_{\text{sr}}/v$ ) / średni interwał między zgłoszeniami ( $a_{\text{sr}}$ ),

= intensywność popytu ( $b_{\text{sr}}/a_{\text{sr}}$ ) / wydajność procesora ( $v$ ).

Przy  $r > 1$  procesor "nie wyrabia się" – **niestabilność**. Przy  $r < 1$  "wyrabia się".

Co się dzieje przy  $r = 1$ ?

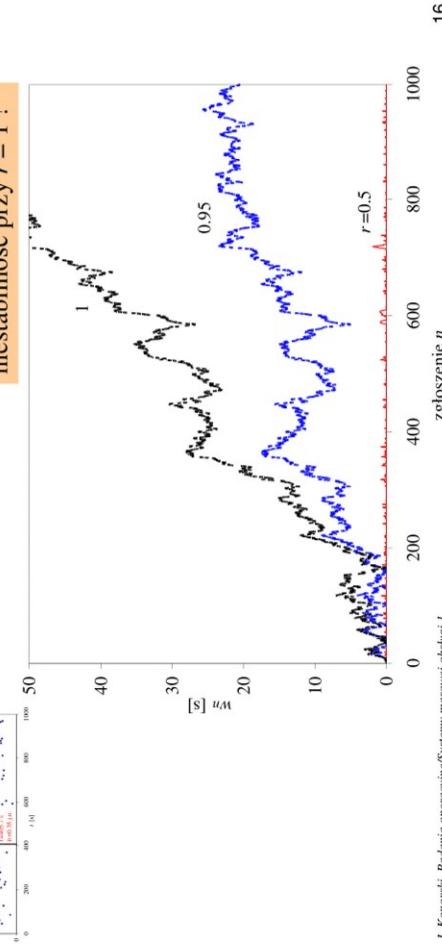
J. Komorski, Badania operacyjne/Systemsy masowej obsługi I

15

## Współczynnik obciążenia – ruch oferowany (2)

ten sam strumień zgłoszeń ( $a_{\text{sr}} = 1$  s,  $b_{\text{sr}} = 1$  j.o.)  
zmniejszanie  $v$

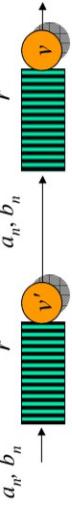
niestabilność przy  $r = 1$ !



J. Komorski, Badania operacyjne/Systemsy masowej obsługi I

## Wpływ prędkości transferu źródło-system

Dodatak zakładalismy natychmiastowy transfer zgłoszeń ze źródła do systemu (dowolnie małe  $a_n$ ). W praktyce prędkość transferu jest ograniczona.



zgłoszenie uważa się za przybyte w chwili rozpoczęcia "wirtualnej obsługi"

Można to modelować jako "wirtualny" wejściowy system kolejkowy z procesorem o wydajności  $v' < \infty$  oraz strumieniem zgłoszeń z  $(b_n)$  i  $(a_n)$ .

W rzeczywistym systemie  $a_n \geq b'_n/v'$ .

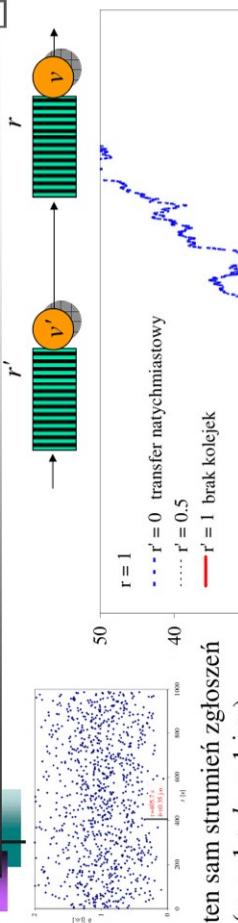
Niech  $r' = (b_{\text{sr}}/a'_{\text{sr}})v'$ ,  $r = (b_{\text{sr}}/a_{\text{sr}})v$ . Oczywiście  $r' = (v/v')r$ .

$$\left. \begin{array}{l} r' = 0 \\ r' = 0.5r \\ r' = r \end{array} \right\} \text{ odpowiada } \left. \begin{array}{l} v' = \infty \text{ (transfer natychmiastowy),} \\ v' = 2v \\ v' = v \text{ (brak kolejek w rzeczywistym systemie).} \end{array} \right.$$

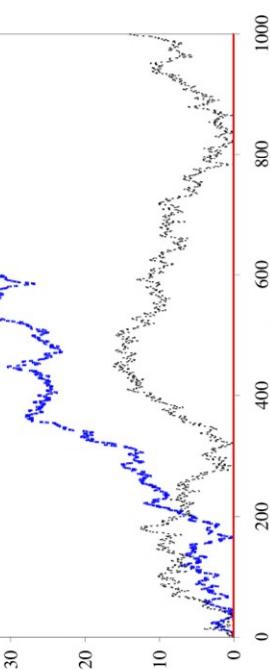
J. Komornicki, Badania operacyjne/Systemy masowej obsługi I

17

## Wpływ prędkości transferu źródło-system (2)



ten sam strumień zgłoszeń  $(a_{\text{sr}} = 1 \text{ s}, b_{\text{sr}} = 1 \text{ j.o.})$



Przy  $r = 1$  lecz skończonej prędkości transferu, system może pozostać stabilny!

J. Komornicki, Badania operacyjne/Systemy masowej obsługi I

18

## W stronę realistycznych modeli



Jaki wpływ na procesy obsługi mają inne charakterystyki strumienia zgłoszeń, zachowanie się zgłoszeń w systemie i zasady obsługi?

- **strumień zgłoszeń**

jakiego mechanizm generacji ( $a_n$ ) i ( $b_n$ )?  
zmienna w czasie? zależność od procesu obsługi? grupowość?  
wymagania  $b_n$  – znane/nieznane w chwili przybycia?

- **pamięć buforowa**

skończona pojemność  $Q$ ?  
ograniczenia dostępności? kiedy odmowa przyjęcia – *drop-tail*?

19

## W stronę realistycznych modeli (2)

• **zachowanie się zgłoszenia w obliczu odmowy przyjęcia / długiej kolejki przepadek?**

ponownie po oczekaniu (*timeout*)?  
eliminacja innych zgłoszeń (*pushout*)?  
nieciepliwość – I, II rodzaju

- **zachowanie się zgłoszenia po rozpoczęciu / zakończeniu obsługi**

warunki opuszczenia kolejki – blokowanie?  
warunki zwolnienia procesora – obsługa także przez inne procesory?  
powrót do kolejki? kiedy? z jakimi zmianami (kreacja zgłoszeń potomnych)?

20



## W stronę realistycznych modeli (3)

- dyscyplina obsługi – co decyduje o kolejności obsługi

chwila przyjęcia (FIFO, LIFO)?

przypadek (RANDOM)?

wielkość wymagania / bieżące zaawansowanie obsługi (SJF, LASF)?

karuzela (RR)?

dodatkowe wymagania zgłoszeń – progi nieciępliwości?

zewnętrzna klasyfikacja zgłoszeń – priorityty, sprawiedliwość (HOL, WFQ)?

- procesor – tryb obsługi

zachowawczy (okres zajętości trwa dokładnie  $\Sigma b_i/v_i$ )

czy niezachowawczy (awarie / strumień ta / "wakacje", obsługa poręciami, zaniechania / wznowienia obsługi)?

bez przepłatu czy z przepłotem?

wiele procesorów – dostępność? wymagania zgłoszeń?

obsługa umiejscowiona czy zrównoleglona?

Czy uda się stworzyć uniwersalny symulator systemów masowej obsługi?

J. Komordi, Badania operacyjne/ Systemy masowej obsługi I

21



## Stan ustalonny

W stacjonarnym systemie masowej obsługi:

- mechanizm generujący strumień zgłoszeń pozostaje niezmienny w czasie (charakterystyki strumienia zgłoszeń usredniane za długą okres obserwacji nie zależą od długości okresu obserwacji)
- wydajność procesora w pozostaje stała.

W dobrze zaprojektowanym systemie stacjonarnym ( $r < 1$ )

występuje stan **ustalonny**  
(charakterystyki procesu obsługi usredniane za coraz dłuższy okres obserwacji ustalają się).

W szczególności

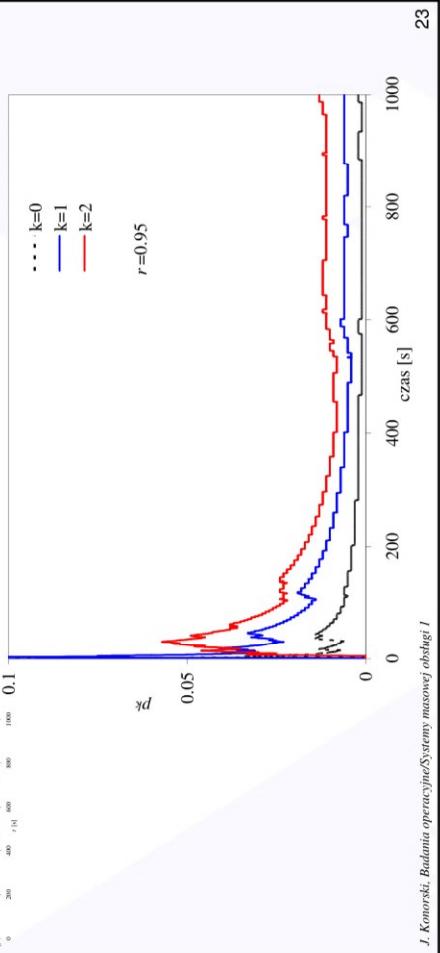
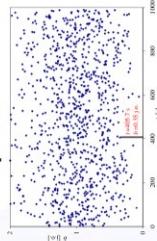
$$\frac{|\{t \leq T \mid N(t) = k\}|}{T} \rightarrow p_k \text{ przy } T \rightarrow \infty$$

J. Komordi, Badania operacyjne/ Systemy masowej obsługi I

22

## Stan ustalony (2)

ten sam strumień zgłoszeń  
 $(a_{\text{sf}} = 1 \text{ s}, b_{\text{sf}} = 1 \text{ j.o.})$



J. Komorski, Badania operacyjne i Systemy masowej obslugi I

## Stan ustalony (3)

Obserwacja  $N(t)$  – różne punkty widzenia

Przy  $n_0 \rightarrow \infty$ :

$$\frac{\#\{n \leq n_0 \mid N(t_n) = k\}}{n_0} \xrightarrow{\text{przypadkowy obserwator}} p_k$$

$$\frac{\#\{n \leq n_0 \mid N(t_n^+ - 0) = k\}}{n_0} \xrightarrow{t} p_k^+$$

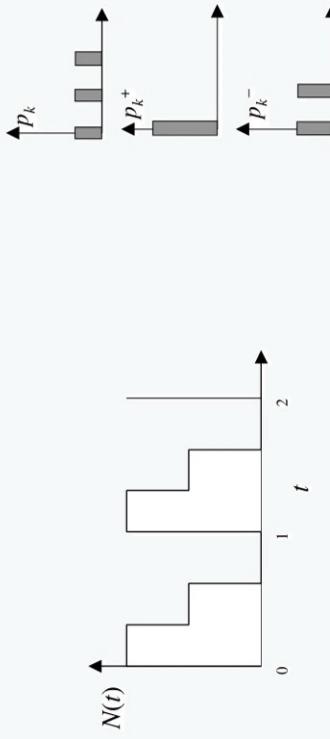
$$\frac{\#\{n \leq n_0 \mid N(t_n^- + 0) = k\}}{n_0} \xrightarrow{t} p_k^-$$

$n_0$  zgłoszenie opuszczające system

J. Komorski, Badania operacyjne i Systemy masowej obslugi I

### Stan ustalony (4)

Uwaga, "praktycy":

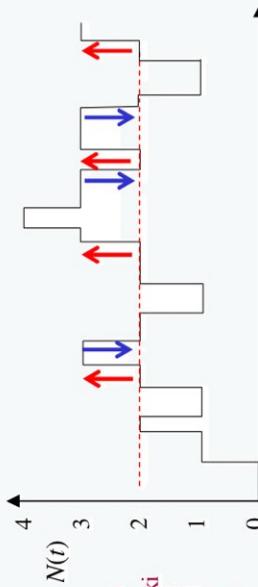


J. Komorski, Badania operacyjne/Sytemy masowej obsługi I

25

### Stan ustalony (5)

Argument skoków  $\pm 1$ : jeżeli  $N(t)$  zmienia się tylko o  $\pm 1$ , to  $p_k^+ \equiv p_k^-$



Różnica między liczbą skoków w góre i w dół = długość kolejki ( $< \infty$  z prawd. 1), zaś różnica między #↑ i #↓ jest  $\leq 1$  w dowolnym przedziale czasu!

26

Uwaga: Jeżeli możliwe są odmowy przyjęcia, to liczą się tylko zgłoszenia przyjęte.  
Np. gdy  $Q < \infty$  (odmowy wskutek przepełnienia pamięci buforowej), to argument skoków  $\pm 1$  daje

$$\frac{p_k^+}{1 - p_Q^+} = p_k^-, \quad k = 0, \dots, Q-1$$

J. Komorski, Badania operacyjne/Sytemy masowej obsługi I



## Stan ustalony (6)

Istotne kryteria oceny systemu:

$$P_1^+ + \dots + P_{Q-1}^+$$

$$L = p_Q^+$$

frakcja zgłoszeń utraconych wskutek przepełnienia pamięci buforowej (bez ponownie)

$$1 - p_0$$

współczynnik zajętości (wykorzystania) procesora  
 $= 1 - \text{współczynnik bezczynności}$

$$N_{\text{sr}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T N(t) dt = \lim_{n_0 \rightarrow \infty} \frac{1}{n_0} \sum_{n=1}^{n_0} N(t_n)$$

średnia liczba zgłoszeń w systemie

$$\frac{W_{\text{sr}}}{d_{\text{sr}}} = \frac{W_{\text{sr}}}{\tau_{\text{sr}}} \cdot \frac{\tau_{\text{sr}}}{v}$$

średnie opóźnienie buforowania

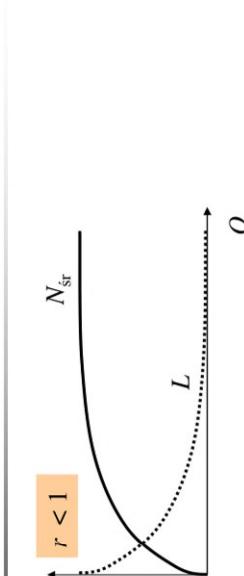
...systemowe

J. Komorski, Badania operacyjne/Systems masowej obsługi I

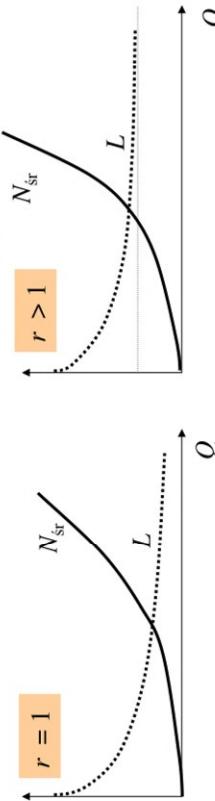
27



## Stan ustalony (7)



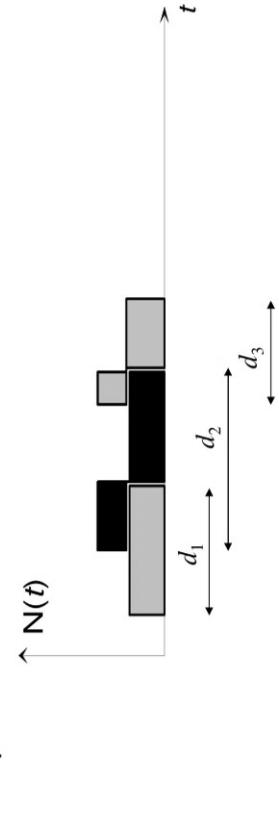
+ systematyczne narastanie kolejki w czasie



J. Komorski, Badania operacyjne/Systems masowej obsługi I

28

### Przepływy



$$\text{przy } n, T \rightarrow \infty : \int_0^T N(t) dt = d_1 + \dots + d_n$$

$$(T/a_{sr})(1-L)d_{sr}$$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

29

### Przepływy (2)

$$N_{sr} = \frac{1-L}{a_{sr}} d_{sr}$$

średnia liczba zgłoszeń w systemie =  
(średnia przepływność systemu) × (średni czas przebywania w systemie)

średnia populacja = (średnia cyrkulacja) × (średni czas życia)

Obowiązuje dla dowolnego fragmentu systemu:

- procesor:  $1 - p_0 = \frac{1-L}{a_{sr}} \tau_{sr} = (1-L)r$  **równanie ciągłości przepływu**
- pamięć buforowa:  $N_{sr} - (1 - p_0) = \frac{1-L}{a_{sr}} w_{sr}$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

30



## Teoria kolejek straci znaczenie?



W przeszłości wzrost  $v$  (a więc spadek  $r$  do zera) wyeliminuje kolejki?

komunikacja transatlantyczka	10 rejsów x 1000 pasażerów/20 dni	1000 lotów x 250 pasażerów/dzień	x500
handel detaliczny	10 sklepów x 100 klientów/dzień	100 sklepów x 10000 klientów/dzień	x1000
iącze dostępowe	10 kb/s	100 Mb/s	x10 000
moc obliczeniowa	\$15m/GFLOPS (1984) <a href="http://en.wikipedia.org/wiki/FLOPS">http://en.wikipedia.org/wiki/FLOPS</a>	\$0.5/GFLOPS (2007)	x30000000

Mimo wszystko odpowiedź: Nie.

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

31



## Teoria kolejek straci znaczenie? (2)

Po pierwsze, samoloty na trasach transatlantycznych, kasły w supermarketach, łącza internetowe, centra komputerowe są zatłoczzone jak nigdy.

Popyt na obsługę rośnie w ślad za podażą –  $r$  wcale nie spada!

32

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

## Teoria kolejek straci znaczenie? (3)

Po drugie, skąd właściwie biorą się kolejki?

Nie tylko z powodu  $v < \infty$ , lecz przede wszystkim z powodu zmienności  $a_n$  (arytmia strumienia zgłoszeń) oraz  $b_n$  (kapryśne wymagania zgłoszeń!).

Arytmia powoduje, że chwilowe obciążenie systemu waha się od 0 do  $\infty$ . Eliminacja choćby tylko okazjonalnych kolejek wymaga procesora z  $v = \infty$ .

Przy  $b_{sr}/a_{sr} < \infty$  daje to  $r = 0$  tj. zerowe wykorzystanie procesora!!!

Bardzo nieekonomiczne, niezależnie od nieprzewidywalnych postępów w technologii i zarządzaniu.

Co jest ekonomiczne? Utrzymywanie  $r < 1$ , tj.  $v > b_{sr}/a_{sr}$ , ale nie bardzo!

Oznacza to dopuszczenie do powstawania kolejek.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

33

## Komentarze do teorii kolejek



Teoria kolejek to analiza matematyczna. Lepszy byłby prototyp / symulator?

- wiarygodne oceny kłopotliwych charakterystyk  
**zdarzenia rzadkie – przekroczenie progu długosci kolejki,**  
dług okres zajętości – czy to już niestabilność?
- jakościowy (a nie swoisty względem określonych ustnień) wpływ  
parametrów modelu na charakterystyki procesów obsługi – **oszczędza**  
**wielu niepotrzebnych eksperymentów**
- ! - uniwersalny (jakościowo, a często i ilościowo) charakter wyników  
dla nieskomplikowanych modeli systemów masowej obsługi

Wbrew pozorom analiza matematyczna jest bardzo kosztowna.  
Oplaca się, gdy udziela odpowiedzi, o które inaczej byłoby trudno.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

34

## Komentarze do teorii kolejek (2)



Agner Krarup Erlang (1878-1929)

duński matematyk i inżynier, pierwszy zrozumiał, że nowoczesna telefonia nie może się obejść bez matematyki jednostka ruchu telekomunikacyjnego = 1 *erlang*

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

35



## Komentarze do teorii kolejek (3)



Jak teoria kolejek "ma się do" analizy:

### szeregowania zadań

tam: optymalny harmonogram dla ustalonego zbioru zadań  
tutaj: ciągły nieprzewidywalny strumień przybywających zadań

### procesów współbieżących

tam: deterministyczna analiza konkretnych scenariuszy zdarzeń  
tutaj: w dużych populacjach zdarzeń (obsługa *masowa*...) widać tylko charakterystyki statystyczne

### procesów przypadkowych

wykorzystuje podobny aparat obliczeniowy  
(proces obsługi = inercyjne przekształcenie strumienia zgłoszeń)

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

36

# Badania operacyjne

## Systemy masowej obsługi

### 2. Modele i charakterystyki stochastyczne

Jerzy Konorski  
jekon@eti.pg.gda.pl

### Zmienne losowe i strumienie losowe

Dla danego scenariusza zgłoszeń można wyznaczyć  $a_{sr}$ ,  $b_{sr}$  i inne charakterystyki. Jednak potrzeba modelu generacji nieprzewidywalnych ( $a_n$ ) i ( $b_n$ ).

Dobrym modelem zjawisk nieprzewidywalnych są **zmienne losowe** – przyjmują konkretne wartości z pewnymi prawdopodobieństwami, tj. posiadają pewne rozkłady prawdopodobieństwa. W ramach powszechnego w nauce i technice podejścia bayesowskiego przyjmujemy te rozkłady za znane / dające się wyliczyć.

Dalej interesować się będziemy wyłącznie **strumieniami losowymi**. Dlaczego niesłoszymy nie?

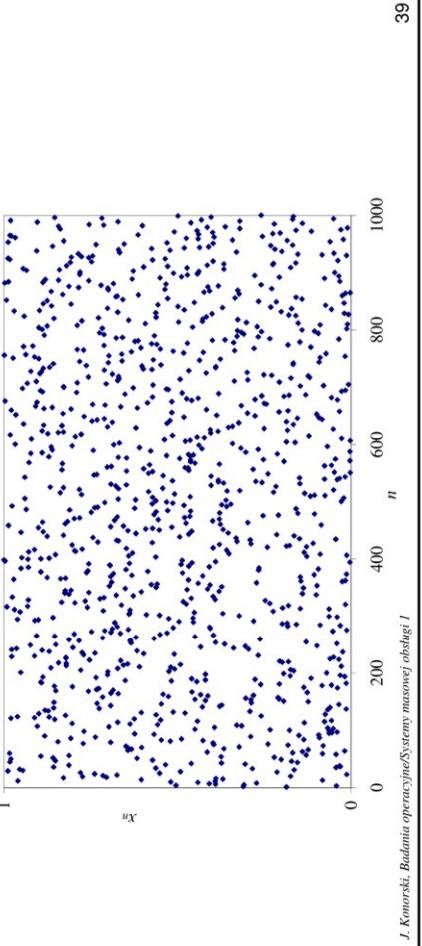
- praktycznie nie występują,
  - nie niosą informacji, nie stanowią problemu projektowego (!)
- Przedmiotem dalszych rozważań będą więc **stochastyczne** modele i charakterystyki systemów masowej obsługi.



## Zmienne losowe i strumienie losowe (2)

Przykładem zmiennej losowej może być liczba zwracana przez random (o ile nie znamy algorytmu generacji liczb pseudolosowych). Posiada ona o rozkład równomierny w przedziale  $[0,1]$ .

Pierwsze 1000 realizacji (zaobserwowanych wartości):

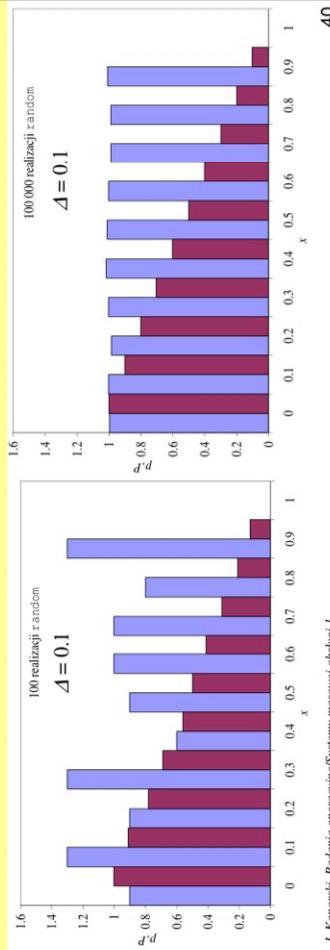


## Rozkłady empiryczne

Na podstawie realizacji  $x_1, \dots, x_N$  tworzymy **histogram**:

- dzielimy zakres możliwych реализаций na przedziały o szerokości  $\Delta$ ,
- zliczamy realizacje w  $i$ -tym przedziale,  $k_i = \#\{n \mid i\Delta \leq x_n \leq (i+1)\Delta\}$
- dla argumentu  $i\Delta$  rysujemy słupek o szerokości  $\Delta$  i wysokości  $p_i = (k_i/N)/\Delta$ ,

**Histogram kumulacyjny** powstaje analogicznie, stupki o wysokości  $P_i = \sum_{j \geq i} p_j$ .

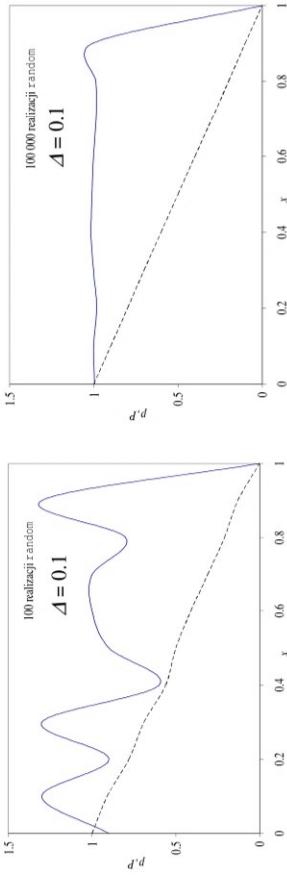


J. Komorski, Badania operacyjne/Sytemy maszynowe obslugi I

40

## Rozkłady empiryczne (2)

...albo bardziej czytelnie, przy pomocy linii ciągły



$$\text{Wartość średnia: } x_{\text{śr}} = \frac{x_1 + \dots + x_N}{N}$$

$$\text{Odchylenie standardowe (rozrzut wokół średniej): } \sigma_x = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x_n - x_{\text{śr}})^2}$$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

41

## Rozkłady teoretyczne

**Gęstość prawdopodobieństwa i komplementarna dystrybuanta**  
– histogramy, jakie otrzymaliśmy w granicy przy  $N \rightarrow \infty$  i  $\Delta \rightarrow 0$ .

$$\text{Gęstość prawdopodobieństwa: } p(x) = \lim_{\Delta \rightarrow 0} \frac{\Pr[x \leq X < x + \Delta]}{\Delta}$$

$$\text{Dla dowolnych } x' \text{ i } x'': \Pr[x' \leq X < x''] = \int_{x'}^{x''} p(x) dx,$$

$$\text{Komplementarna dystrybuanta: } P(x) = \int_x^{\infty} p(y) dy$$

$$x_{\text{śr}} = \int_{-\infty}^{\infty} xp(x) dx, \quad \sigma_x = \sqrt{\int_{-\infty}^{\infty} (x - x_{\text{śr}})^2 p(x) dx}$$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

42

## Rozkłady teoretyczne (2)

W wielu problemach inżynierskich występuje tzw. rozkład Weibulla:

$$P(x) = e^{-\lambda x^\theta}, \quad p(x) = \lambda \theta x^{\theta-1} e^{-\lambda x^\theta}, \quad x \geq 0$$

$\lambda$  – parametr skali,  
 $\theta$  – parametr kształtu (= 1: rozkład wykładniczy).

$$x_{\text{sr}} = \int_0^\theta \sqrt{\frac{y}{\lambda}} e^{-y} dy$$

$$\sigma_x = \sqrt{\int_0^\infty \frac{y^2}{\lambda^2} e^{-y} dy - x_{\text{sr}}^2}$$

J. Komorski, Badania operacyjne i Systemy masowej obsługi I

43

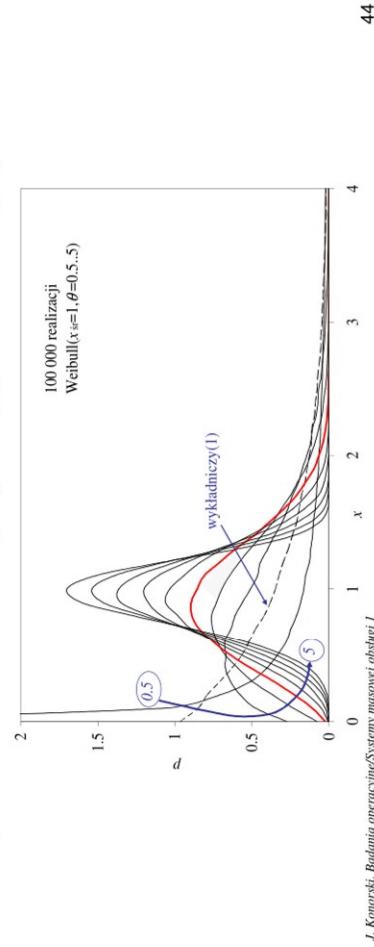
## Sztuczna generacja strumieni zgłoszeń



Mamy generator r. random liczb pseudolosowych ( $z_n$ ).  
 Jak generować liczby pseudolosowe ( $x_n$ ) o dowolnej funkcji  $P(x)$ ?

$$x_n = \theta \sqrt{-\frac{\ln z_n}{\lambda}}$$

(Izw. metoda odwracania dystrybuanty; jest też wiele innych).



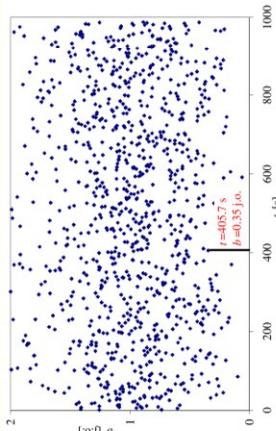
J. Komorski, Badania operacyjne i Systemy masowej obsługi I

## Sztuczna generacja strumieni zgłoszeń (2)

Wprowadźmy zmienné losowe:

$A$  – interwał (realizacje  $a_n$ ),

$B$  – wymaganie (realizacje  $b_n$ )



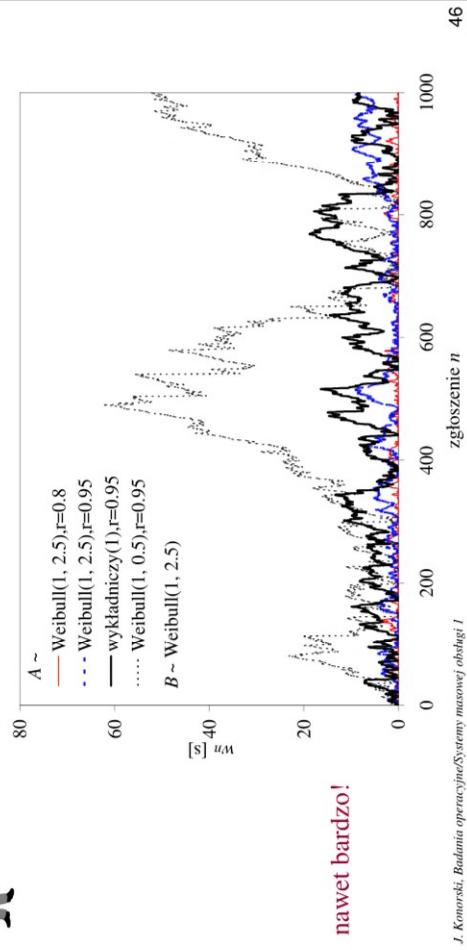
Znany nam już strumień został wygenerowany jako  $A, B \sim \text{Weibull}(1, 2.5)$ .

J. Komorski, Badania operacyjne/Sytremy masowej obslugi I

45

## Wpływ kształtu rozkładu $A, B$

Widzieliśmy, że przebieg procesu obsługi silnie zależy od wartości średnich  $a_{\text{sr}}, b_{\text{sr}}$ . Czy także od kształtu rozkładów  $A$  i  $B$ ?



J. Komorski, Badania operacyjne/Sytremy masowej obslugi I

nawet bardzo!

46

## Notacja Kendall'a

A/B/S/Q/J ...

Przykłady:

M/M/1	$S - \# \text{procesorów}$
M/D/1/10	$Q - \text{pojemność pamięci buforowej}$
$E_3/H_2/5/10$	$J - \text{rozmiar populacji źródeł zgłoszeń}$
M/M/5/20	Typy rozkładów:
M/G/S	M – wykładniczy
G/G/S	D – deterministyczny
M <sup>X</sup> /G/5	$E_k - \text{Erlanga rzędu } k$
M/MPP/G/1	$H_k - \text{hiperwykładniczy rzędu } k$
D+M/M/1/20	G – ogólny
M/D/1/P*Q	
M/G/1 LIFO, HOL, ...	

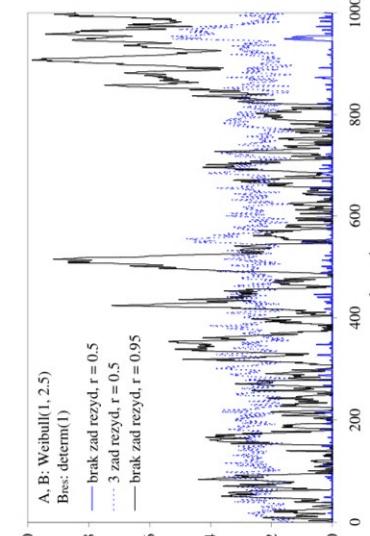


Jaki jest wpływ innych czynników – S, Q, zachowana zgłoszeń, zasad obsługi?

J. Komorski, Badania operacyjne/Sytemy masowej obsługi I

47

## Zadania rezydентne



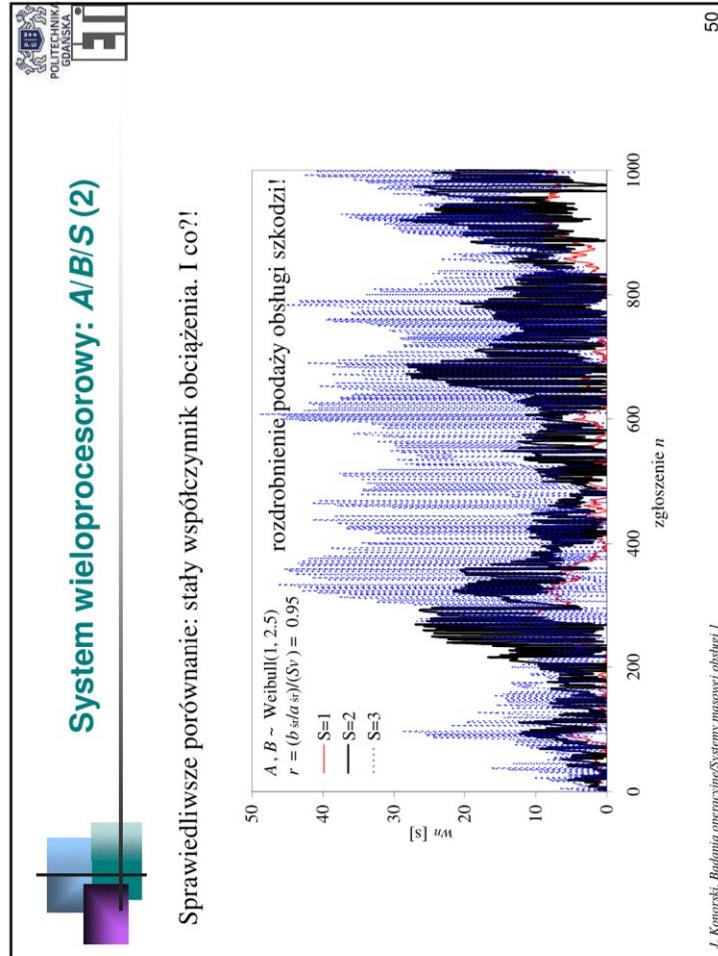
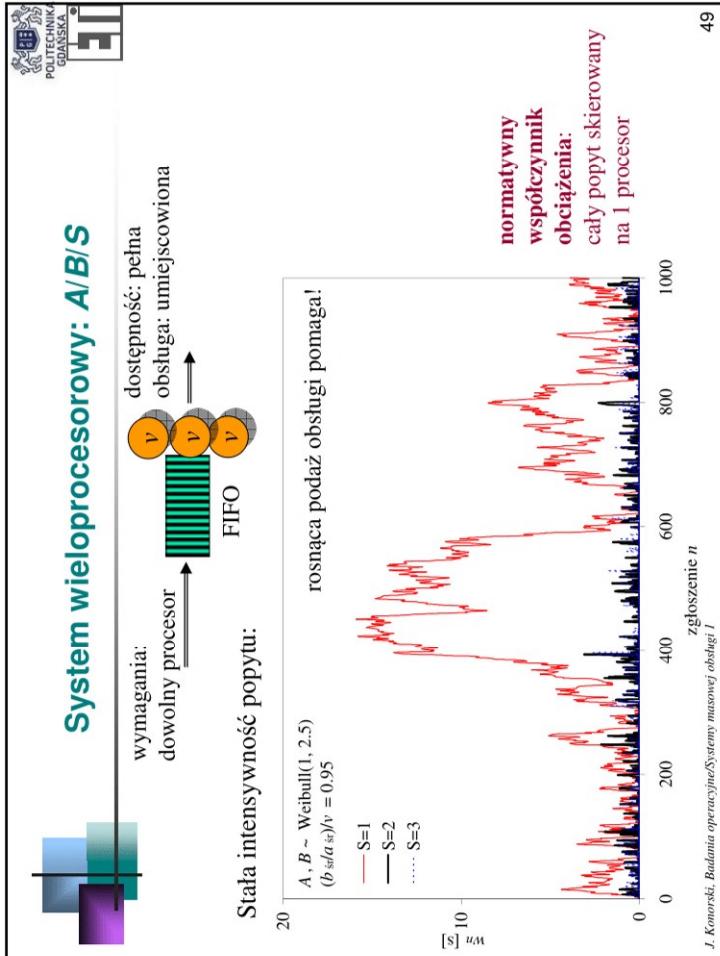
Przyczyną wzrostu opóźnień jest wzrost r?

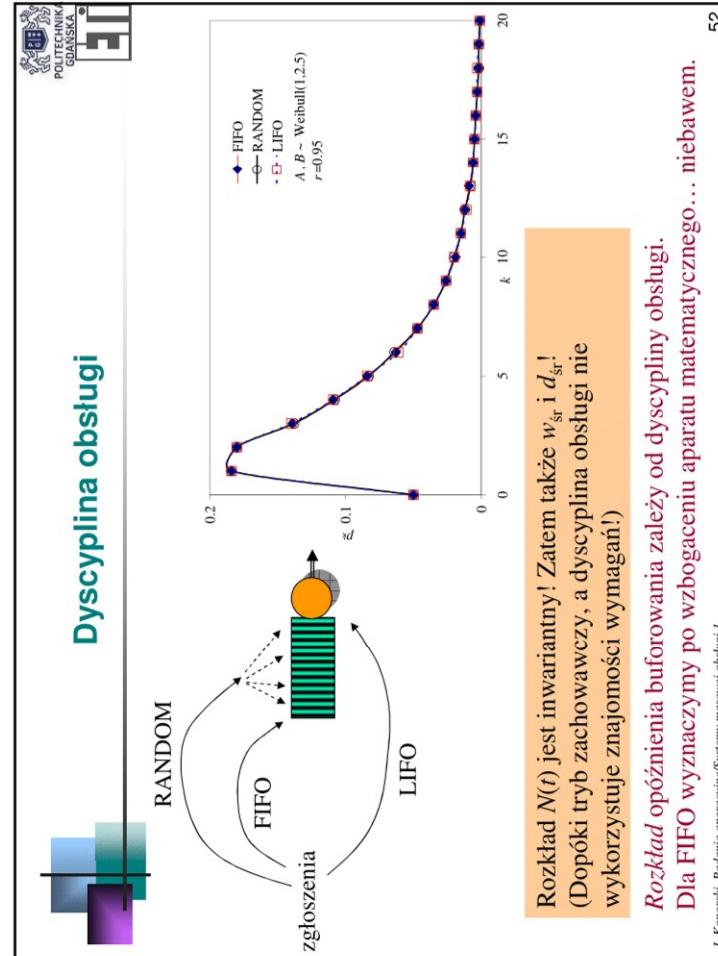
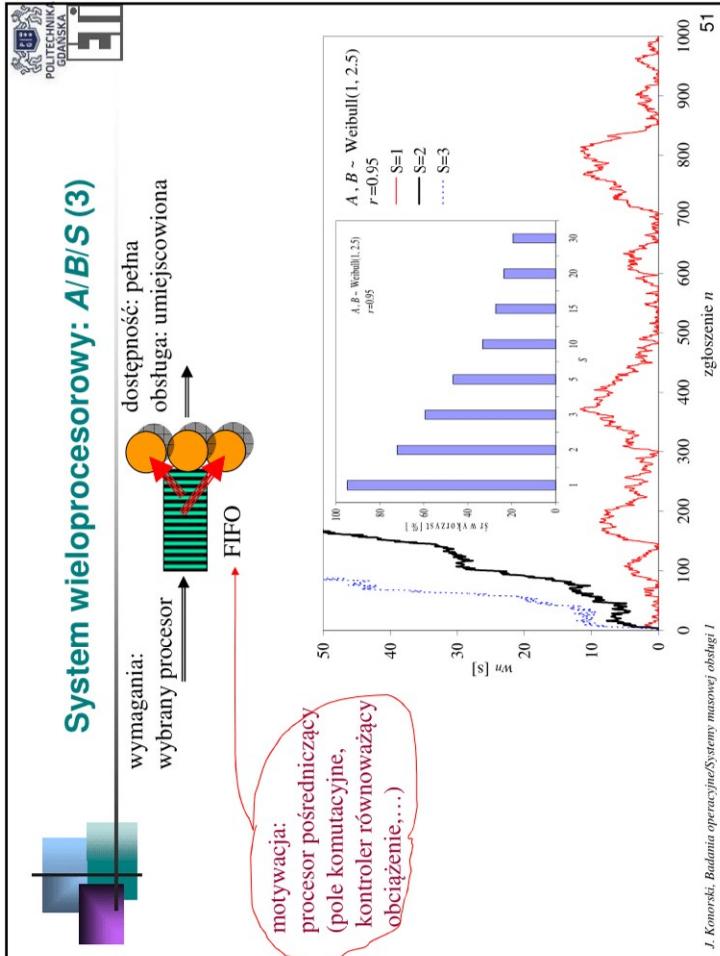
$$\frac{b_{\text{sr}}}{a_{\text{sr}} V} + \frac{\# \text{rezyd} \cdot b_{\text{sr}}}{a_{\text{sr}} V} \approx 0.5 + \frac{3 \cdot 1}{3.5 \cdot 2} \approx 0.95$$

Porównaj wykresy dla  $r = 0.95$  – jest też wpływ jakościowy.

J. Komorski, Badania operacyjne/Sytemy masowej obsługi I

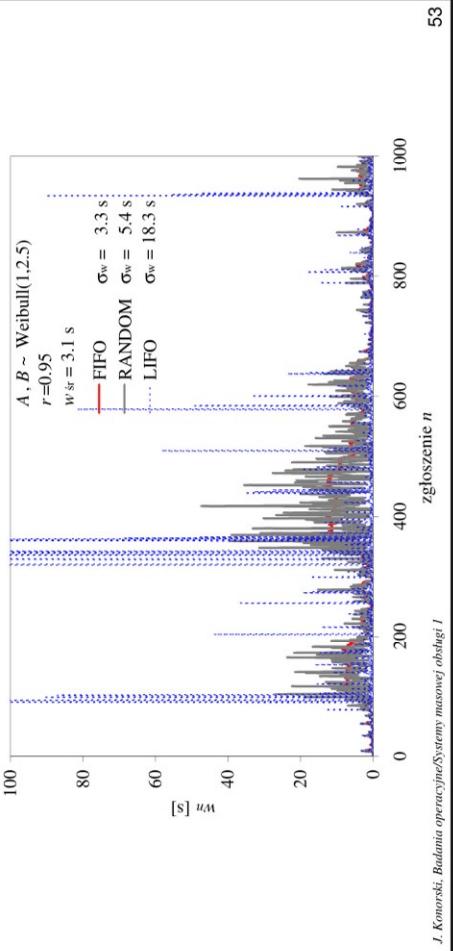
48





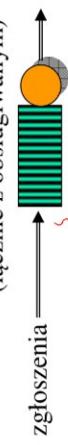
## Dyscyplina obsługi (2)

Która dyscyplina powoduje największe  $w_{\text{sr}}$ ? A która największe  $\sigma_w$ ?



## Skończona pojemność bufora: $A/B/1/Q$

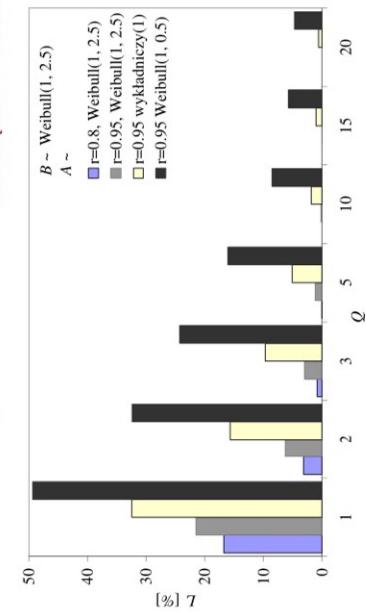
pojemność  $Q$  zgłoszeń  
(łącznie z obslugiwany)



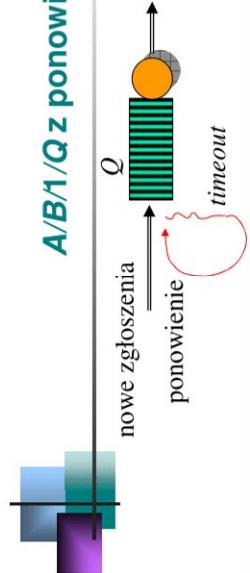
Motywacja:

- koszt pamięci?!
- zarządzanie, opóźnienia.

Mniejsze opóźnienia buforowania  
za cenę strat wskutek przepchnienia.



## A/B/1/Q z ponowieniami



W bardziej realistycznym modelu zgłoszenie odrzucone nie jest utracone, lecz oczekuje pewien czas i przybywa znów (**ponowienie**).

- zajęty numer telefoniczny ("spróbuj zadzwonić później")
- chwilowo niedostępny serwer/sieć GSM, ...

Jako  $w_n$  przyjmujemy czas od *pierwszego zgłoszenia do rozpoczęcia obsługi*. Odzwierciedla to jednocześnie opóźnienia i przepełnienia pamięci buforowej.

Prosty model:

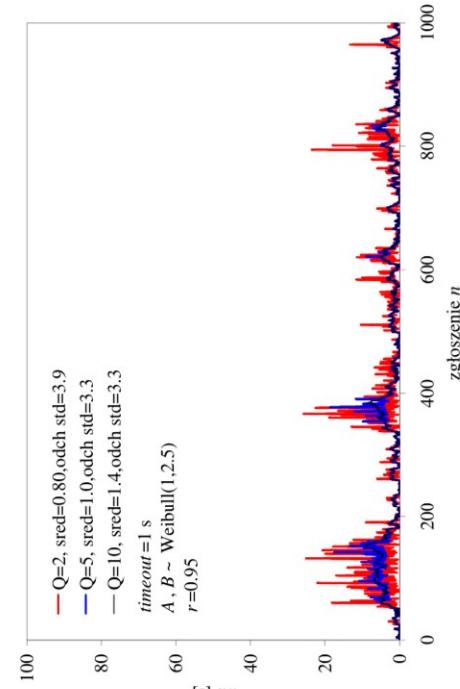
- stały czas oczekania (*timeout*)
- nieograniczona liczba ponowień.



*J. Komorski, Badania operacyjne/Syttemy masowej obslugi I*

55

## A/B/1/Q z ponowieniami (2)



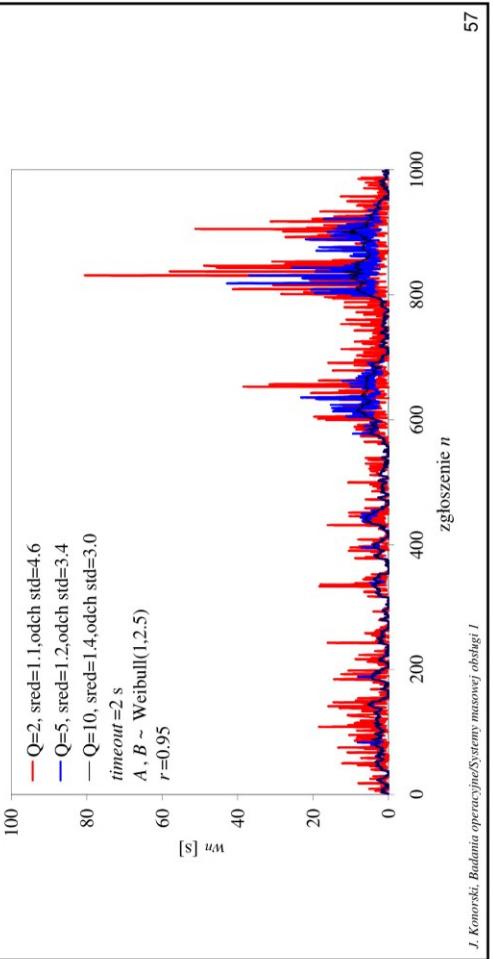
Zmniejszanie  $Q$  polepsza średnią, pogarsza rozrzut i "ogony" rozkładu opóźnień.

*J. Komorski, Badania operacyjne/Syttemy masowej obslugi I*

56

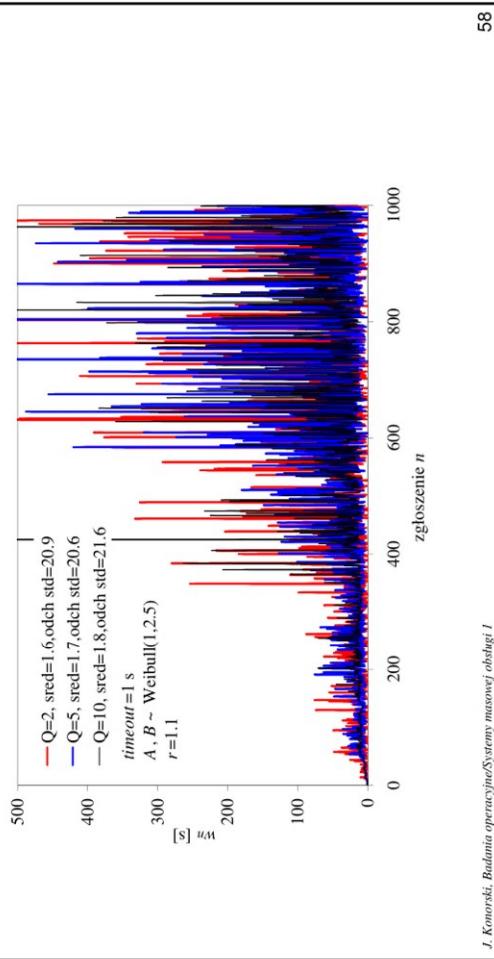
### A/B/1/Q z ponowieniami (3)

Staje się to bardziej widoczne przy "cierpliwszych" zgłoszeniach:



### A/B/1/Q z ponowieniami (4)

Przy nieco większym obciążeniu dramatycznie widoczne są "ogony":





## Ważniejsze modele strumieni zgłoszeń



- Weibulla strumienie odnowy;
  - Bernoulliego ( $a_n$ ) jest ciągiem realizacji niezależnych zmiennych losowych o wspólnym rozkładzie (**iid** – *independent, identically distributed*)
  - Erlanga
  - gamma
  - ...
- bardziej skomplikowane
- zmienne w czasie rozkłady A i B, MMPP, BMAP, fraktałny proces Browna, ...
- modelują niestacjonarność, zależność od procesu obsługi, grupowość, wewnętrzna korelacja, ...

J. Komorski, Badania operacyjne/Sytsemy masowej obsługi I

59



## Wewnętrzna korelacja



Czy o przebiegu procesu kolejowego przesądzają same rozkłady A i B, czy też rolę odgrywa **wewnętrzna korelacja (autokorelacja)** w ciągu ( $a_n$ )?

$$\text{corr}_a(l) = \frac{1}{\sigma_a^2} \frac{1}{M} \sum_{n=1}^M (a_n - a_{\text{sr}})(a_{n+l} - a_{\text{sr}}), \quad M \rightarrow \infty, l = 0, 1, 2, \dots$$

**funkcja autokorelacji** = jak bardzo skorelowane są ze sobą interwaly odległe o  $l$  zgłoszeń; korelacja zwykle zmala przy rosnącym  $l$

W strumieniu odnowy brak jest wewnętrznej korelacji (~ "biały szum"):

$$\text{corr}_a(l) = \begin{cases} 1, & \text{gdy } l = 0 \\ 0, & \text{gdy } l \neq 0 \end{cases}$$

J. Komorski, Badania operacyjne/Sytsemy masowej obsługi I

60

## Wewnętrzna korelacja (2)

### Eksperyment 1

Generujemy ciągi  $(a_n)$  i  $(b_n)$  zgodnie z Weibull(1, 2.5) metodą odwracania dystrybuanty, wykorzystując kolejne liczby pseudolosowe. Tak powstający strumień odnowy kierujemy do systemu masowej obsługi oslugi z  $r = 0.95$ .

Obserwujemy proces kolejkowy  $(w_n)$ .

Następnie **tasujemy**, tj. przypadkowo permutujemy wygenerowany ciąg  $(a_n)$ , bierzemy ten sam ciąg  $(b_n)$  i powtarzamy obserwację  $(w_n)$ .

## Wewnętrzna korelacja (3)

$A \sim$

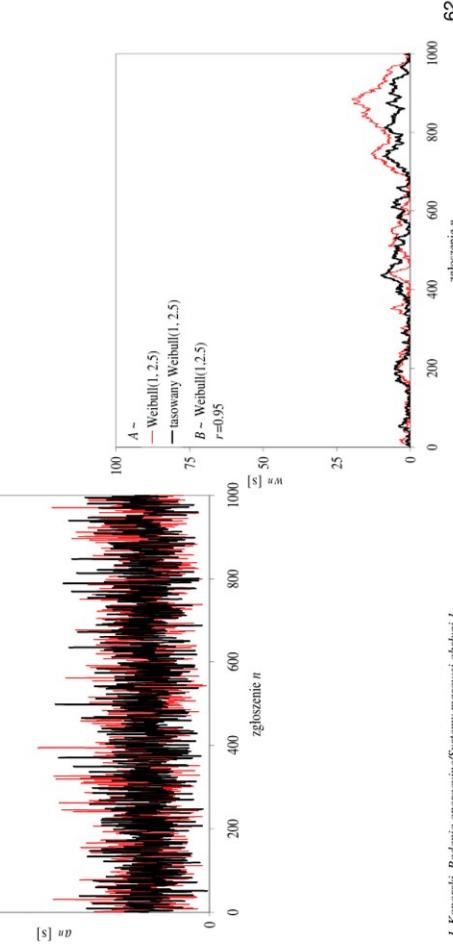
— Weibull(1, 2.5)

— tasowany Weibull(1, 2.5)

$B \sim$

Weibull(1, 2.5)

$r=0.95$



## Wewnętrzna korelacja (4)

### Eksperyment 2

Generujemy ciąg  $(a_n)$  metodą odwracania dystrybutanty,  $A \sim \text{Weibull}(1, 2.5)$ ,

- wykorzystując kolejne liczby pseudolosowe – strumień odnowy
- powtarzając interwał liczbę razy wylosowaną z rozkładu Weibull(1, 0.5)

W obu wariantach rozkład interwala jest taki sam. Jednak drugi wariant daje  $(a_n)$  z dalekosieczną wewnętrzną korelacją – tzw. **strumień samopodobny**.

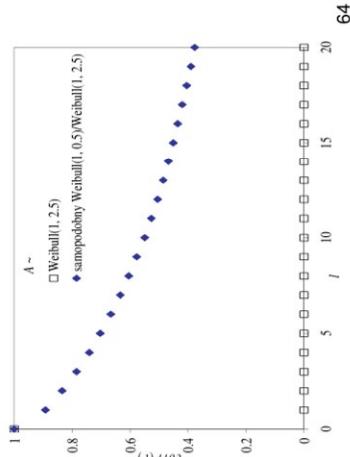
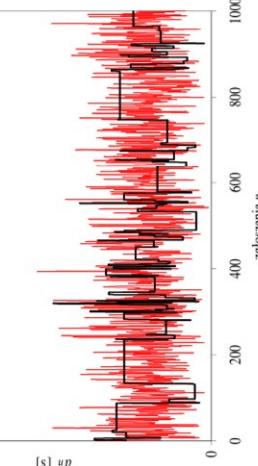
Bierzemy ten sam ciąg  $(b_n)$  i znów kierujemy tak powstały strumień zgłoszeń do systemu masowej obsługi, w którym  $r = 0.95$ .

## Wewnętrzna korelacja (5)



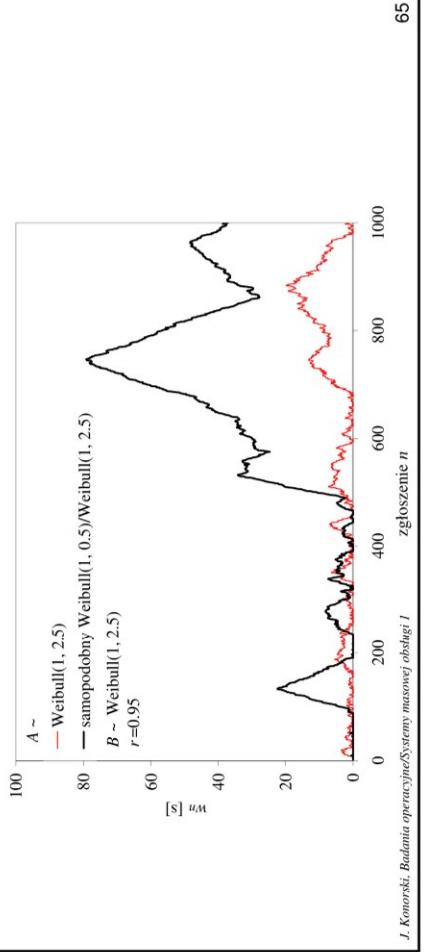
$A \sim$   
— Weibull(1, 2.5)

— samopodobny Weibull(1, 0.5)/Weibull(1, 2.5)



## Wewnętrzna korelacja (6)

Porównanie procesów kolejkowych dla strumienia odnowy i samopodobnego wypada dramatycznie...

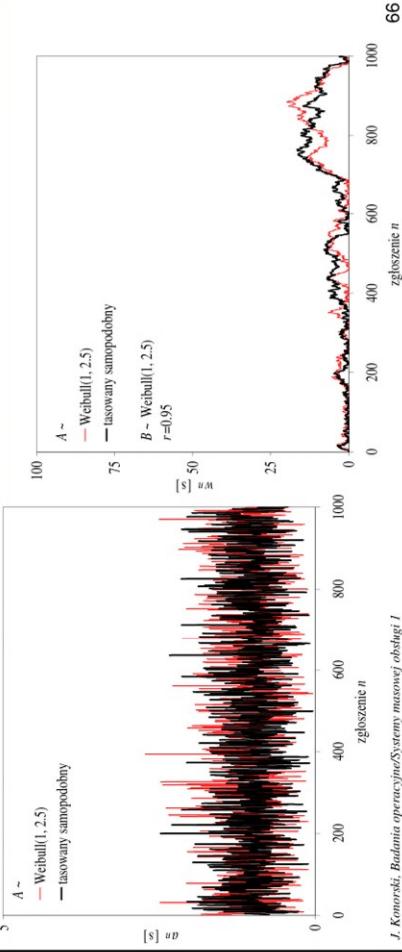


## Wewnętrzna korelacja (7)

### Eksperyment 3

Tasujemy ciąg  $(a_n)$  we wcześniejszy wygenerowanym strumieniu samopodobnym, co niszczy wewnętrzną korelację, lecz zachowuje rozkłady A i B.

Bierzemy ten sam ciąg  $(b_n)$  i porównujemy ze strumieniem odnowy...





## Wewnętrzna korelacja (8)

Wnioski:

- Tasowanie strumienia odnowy nie zmienia (braku) wewnętrznej korelacji, nie wpływa też istotnie na przebieg procesu kolejkowego – można się o tym przekonać np. konstrując histogram opóźnien buforowania. W przypadku strumienia odnowy rozkłady A i B stanowią wystarczającą informację dla przewidywania własności procesu kolejkowego.
- Obecność wewnętrznej korelacji w strumieniu zgłoszeń może wyraźnie pogarszać przebieg procesu kolejkowego – obserwujemy to na przykładzie strumienia samopodobnego (z dalekosieczną wewnętrzną korelacją). Tutaj o przebiegu i własnościach procesu kolejkowego decyduje głównie postać funkcji autokorelacji ciągu  $(a_n)$ .
- Znaczenie wewnętrznej korelacji staje się widoczne po jej usunięciu poprzez tasowanie.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

67



## Strumień odnowy: interwał resztkowy

Z wyjątkiem szczególnych zastosowań (rytm serca, wylewy Niliu, ruch WWW,...) strumienie odnowy stanowią dobry model rzeczywistych strumieni zgłoszeń. Przez chwilę skupimy uwagę tylko na interwałach.



Jakie typy rozkładów A są spotykane w praktyce?

Naturalnie bardzo różne. Jednak istnieje pewien rozkład o uniwersalnym znaczeniu i pewnych "magicznych" właściwościach.  
Aby je zrozumieć, należy prześledzić problem **interwala resztkowego**.

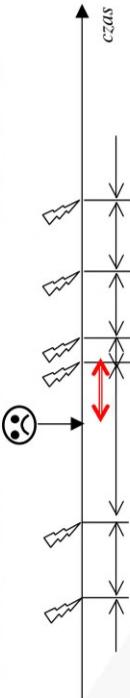
Zdarzenia zachodzą z przypadkowymi interwałami. Wybieramy przypadkową chwilę. Jak długo poczekamy na najbliższe zdarzenie?

Mожет to być problem pasażera autobusu, abonenta dodzwaniającego się do zajętego numeru telefonicznego, nieprzerwającego zgłoszenia VIP i in.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

68

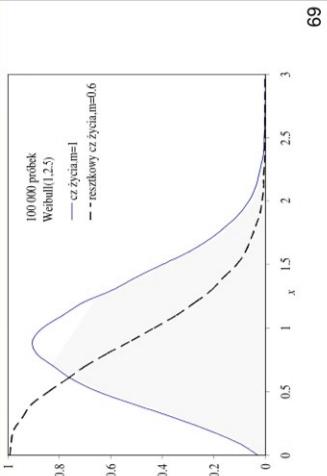
## Strumień odnowy: interwał resztkowy (2)



Autobus przyjeżdża średnio co godzinę. "Statystyczny pasażer" czeka pół godziny?

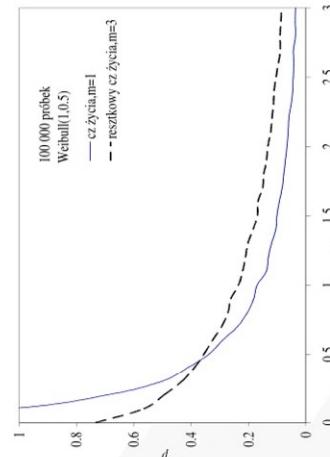
Znamy rozkład interwala między autobusami. Jaki przewidujemy rozkład interwala resztkowego? Przesunięty w stronę mniejszych realizacji?

J. Komorski, Badania operacyjne/Systemy masowej obsługi I



69

## Strumień odnowy: interwał resztkowy (3)



Paradoks resztkowego czasu życia:

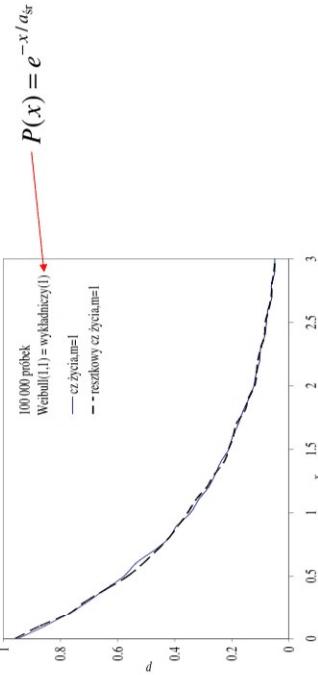
$$\tilde{a}_{\text{sr}} = \frac{a_{\text{sr}}}{2} (1 + c_a^2) \quad c_a = \frac{\sigma_a}{a_{\text{sr}}} - \text{współczynnik rozproszenia rozkładu } A$$

Jeżeli  $A \sim P(x)$ , to  $\tilde{A} \sim \tilde{P}(x) = P(x)/a_{\text{sr}}$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

70

### Strumień odnowy: interwał resztkowy (4)



Rozkład wykładniczy jest bezpamięciowy:

$$\Pr[A < x + \Delta | A \geq x] = \frac{P(x)\Delta}{P(x)} = \frac{\Delta}{a_{sr}} \quad \text{bez względu na } x$$

J. Komorski, Badania operacyjne i Systemy masowej obsługi I

71

### Strumień Poissona



Simeon-Denis Poisson (1781-1840)

matematyk, fizyk i astronom francuski

badał bezpamięciowe procesy stochastyczne – procesy Poissona – dobrze modelujące współczesny ruch telekomunikacyjny i komputerowy

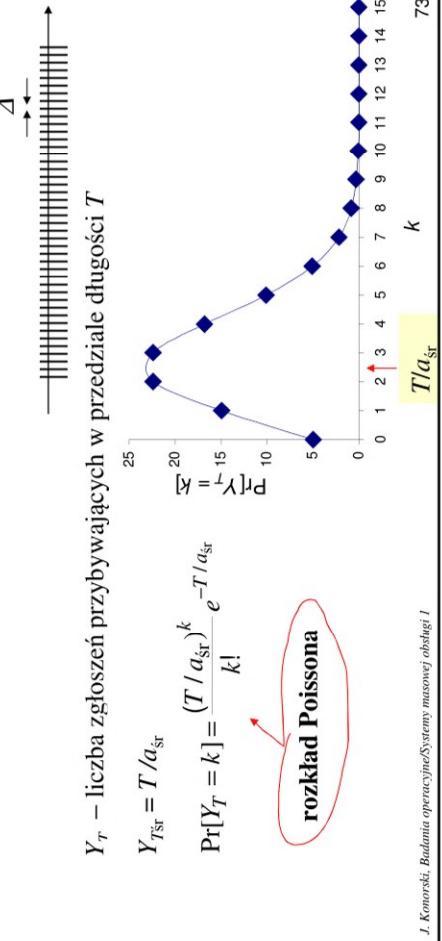
J. Komorski, Badania operacyjne i Systemy masowej obsługi I

72

## Strumień Poissona (2)

W każdej chwili nowe zgłoszenie przybywa ze stałym prawdopodobieństwem.  
W notacji Kendallu: systemy  $M/\dots$

$$\Pr[\text{zgłoszenie w } (t, t + \Delta)] = \Delta a_{\text{sr}} + o(\Delta)$$



## Strumień Poissona (3)

Dla stacjonarnych modeli stochastycznych charakterystyki procesu obsługi mające sens średnich czasowych wyznaczamy z pomocą rachunku prawdopodobieństwa.

Stan ustalony nazywa się wtedy **równowagą statystyczną**.

Polega na ustalaniu się w czasie interesujących nas średnich, np. prawdopodobieństw stanów systemu / utraty zgłoszenia / opóźnienia buforowania / ...

**PASTA (Poisson Arrivals See Time Averages):** dla strumienia Poissona  $p_k^+ \equiv p_k$   
(przybywające zgłoszenia tworzące strumień Poissona "widzą" taki sam rozkład długosci kolejki jak przypadkowy obserwator)

$$\Pr[N(t) = k] \frac{\Delta}{a_{\text{sr}}} = \Pr[N(t) = k]$$

Dowód:  $\Pr[N(t) = k | \text{przybycie w } (t, t+\Delta)] = \frac{\Delta}{a_{\text{sr}}} = \Pr[N(t) = k]$

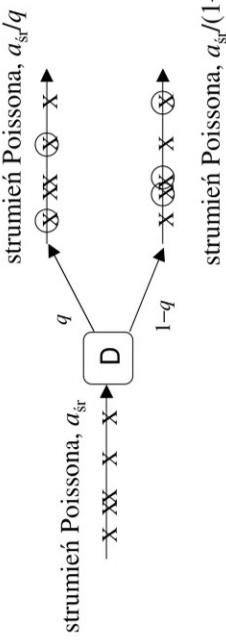
**M/G/S/Q:** straty wskutek przepchnięcia pamięci buforowej  $L = p_Q^+ \equiv p_Q$ .

*J. Komorski, Badania operacyjne i Systemy masowej obsługi I*



## Strumień Poissona (4)

Rozrzedzanie losowe:



Przy rozrzedzaniu losowym strumień Poissona zachowuje swój charakter!

Rozrzedzanie oparte o inne mechanizmy już tej własności nie ma.

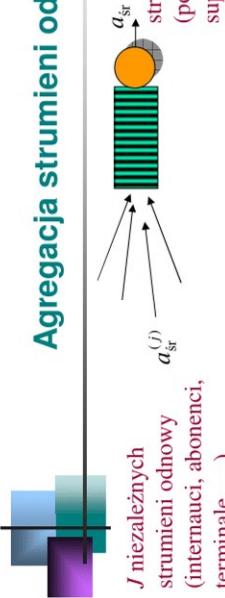
Rozrzedzanie losowe innego typów strumienia – także nie.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

75



## Agregacja strumieni odnowy

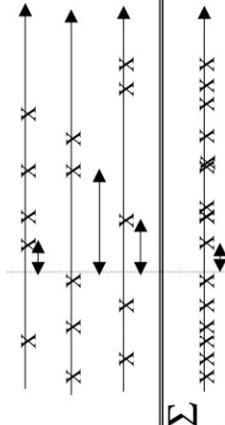


$J$  niezależnych  
strumieni odnowy  
 $a_{sr}^{(j)}$   
(internauci, abonenci,  
terminali, ...)

strumień zagregowany  
(podawany na tące szkieletowe,  
superkomputer / serwer WWW...)

$$\frac{T}{a_{sr}} = \sum_{j=1}^J \frac{T}{a_{sr}^{(j)}} \quad (\text{przy jednakowych strumieniach składowych } a_{sr} = \frac{a_{sr}^{(j)}}{J})$$

Jaki rozkład interwalu "widzi" system? Czy przy najmniej strumień odnowy?  
W ogólności nie. Obliczenia b. trudne.



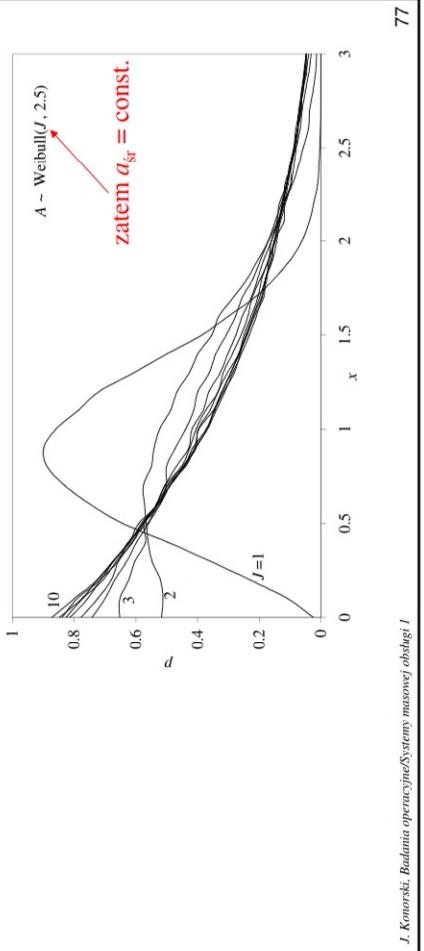
J. Komorski, Badania operacyjne/Systemy masowej obsługi I

76

## Agregacja strumieni odnowy (2)

Przy  $J \rightarrow \infty$ , ale  $a_{\text{sr}} > 0$  (praktyczna sytuacja) oraz przy założeniu, że strumienie składowe są niezależne:

Strumień zagregowany jest strumieniem Poissona (twierdzenie Palma).



## Agregacja strumieni odnowy (3)

Dowód pominiemy :

$$\tilde{A} = \min\{\tilde{A}^{(1)}, \dots, \tilde{A}^{(J)}\}, \text{ zatem } \Pr[\tilde{A} \geq x] = \prod_{j=1}^J \Pr[\tilde{A}^{(j)} \geq x]$$

$J \rightarrow \infty$ , ale  $a_{\text{sr}} > 0$  (praktyczna sytuacja). Czyli  $a_{\text{sr}}^{(j)} \rightarrow \infty$ .

Dla dowolnego skończonego  $x$ :  $x/a_{\text{sr}}^{(j)} \rightarrow 0$  i możemy zaniedbać  $Y_x^{(j)}$   $> 1$ .

$$\Pr[\tilde{A}^{(j)} \geq x] = 1 - \Pr[Y_x^{(j)} > 0] \approx 1 - \Pr[Y_x^{(j)} = 1] \approx 1 - Y_{x, \text{sr}}^{(j)} \approx 1 - \frac{x}{a_{\text{sr}}^{(j)}}$$

Ostatecznie

$$\Pr[\tilde{A} \geq x] \approx \prod_{j=1}^J \left( 1 - \frac{x}{a_{\text{sr}}^{(j)}} \right) \approx \exp\left( - \sum_{j=1}^J \frac{x}{a_{\text{sr}}^{(j)}} \right) = e^{-x/a_{\text{sr}}}$$

gdyż przy  $x << 1$  mamy  $1 - x \approx e^{-x}$

J. Komorski, Badania operacyjne i systemy masowej obsługi I



# Badania operacyjne

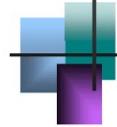
## Systemy masowej obsługi

### 3. Modele markowskie

Jerzy Konorski  
jekon@eti.pg.gda.pl

*J. Konorski, Badania operacyjne/Systemy masowej obsługi I*

79



### Systemy markowskie

Przypomnienie: teoria kolejek zajmuje się problemami masowej obsługi, tj. systemami i procesami, które można obserwować, mierzyć i symulować.

Analiza matematyczna systemów masowej obsługi też jest przydatna, ale tylko wówczas, gdy przynosi *proste i pouczające* wyniki.

Rozpatrzmy model A/B/... Czy charakterystyki procesu obsługi łatwo jest przewidzieć teoretycznie?

Tak – o ile wprowadzimy niezbędne uproszczenia:

- niezbyt drastyczne - związek z rzeczywistością!
- inaczej zarżut "szukania pod latarnią, bo tam jest jasno"
- znaczne – prostota analizy! uniwersalne walory poznawcze!



Przykład: systemy markowskie.

*J. Konorski, Badania operacyjne/Systemy masowej obsługi I*

80

## Systemy markowskie (2)

Mają niezwykłą i niezwykle przydatną w modelowaniu właściwość **Markowa**.

(Która co prawda dzieli z ogromną liczbą występujących w rzeczywistości systemów dynamicznych: technicznych, fizycznych, społecznych, biologicznych, ekonomicznych, ...)

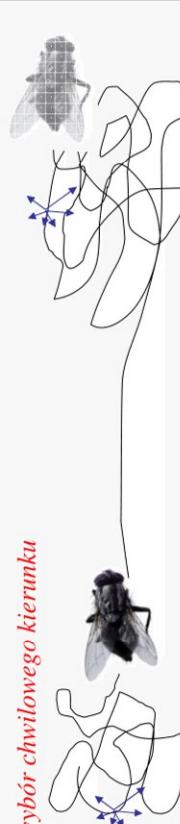
$$stan(t + \Delta) = f[stan(t), \phi]$$

losowe "pobudzenie zewnętrzne" w chwili  $t$   
na ogół zależne od bieżącego stanu,  
jednak niezależne od stanów wcześniejszych

## Systemy markowskie (3)

• Mucha...

*wybór chwilowego kierunku*



• tasowanie talii kart: kolejność kart (*wybór karty do przetwarzania*)

• kapitał gracza / populacja (*bieżąca stopa zysku / przyrost naturalny*)

•  $trend(t + \Delta) = (1 - c) \cdot trend(t) + c \cdot \phi(t)$  (*kolejna obserwacja*)

•  $udzial_w_{rymku}(t + \Delta) = \phi \cdot udzial_w_{rymku}(t)[1 - udzial_w_{rymku}(t)]$   
(*bieżąca jakość zarządzania*)

• topologia Internetu (*liczba nowych sieci i miejsc ich dołączenia*)

## Systemy markowskie (4)



### Andriej Andriejewicz Markow (1856-1922)

matematyk rosyjski

badał procesy przypadkowe o ograniczonej pamięci – *procesy Markowa* – doskonale modelujące wiele zjawisk przyrodniczych i technicznych

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

83



## Systemy markowskie (5)

Są nimi systemy obsługi M/M/...

strumień Poissona,  $a_{sr}$  → wykładniczy rozkład wymagań,  $b_{sr}$  →  $v$



Znaczenie:

- strumień Poissona:
  - twierdzenie Palma
  - rozszerzanie losowe
  - PASTA
  - pesymistyczne oceny charakterystyk procesów obsługi
- wykładniczy rozkład wymagań:
  - rozmowa telefoniczna
  - transfer P2P
  - ...

84

*J. Komorski, Badania operacyjne/Systemy masowej obsługi I*

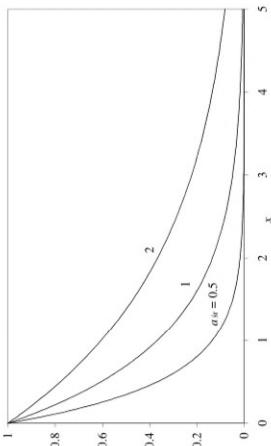
## Systemy markowskie (6)

$v$  – (stala) wydajność procesora

$A \sim M$  = wykładniczy ( $a_{sr}$ ):  $\Pr(A \geq x) = e^{-x/a_{sr}}$

$B \sim M$  = wykładniczy ( $b_{sr}$ ):  $\Pr(B \geq x) = e^{-x/b_{sr}}$

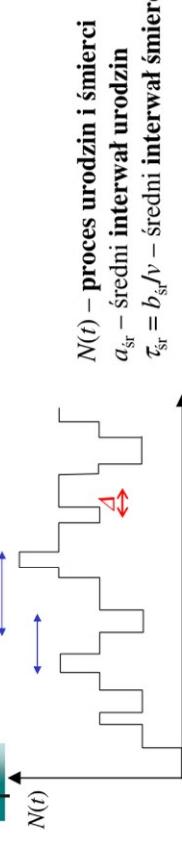
$\Pr[\text{czas obsługi} \geq x] = e^{-x/(b_{sr}/v)}$



J. Komorski, Badania operacyjne/Systemy masowej obsługi I

85

## System M/M/1



Niech  $p_k(t) = \Pr[N(t) = k]$ . Wiemy, że  $\lim_{t \rightarrow \infty} p_k(t) = p_k$  (równowaga statystyczna).  
Podstawą wyznaczenia rozkładu  $(p_k)$  jest wypisanie równań urodzin i śmierci

- Rozkład wykładniczy nie posiada "atomu" w punkcie 0

⇒ Obserwujemy stan  $N(t) = k$ . Co może zdarzyć się między  $t$  a  $t + \Delta t$ ?  
praktycznie tylko: nic / 1 urodziny / 1 śmierć (gdy  $k > 0$ ),  
(przejście jedynie między sąsiednimi stanami)

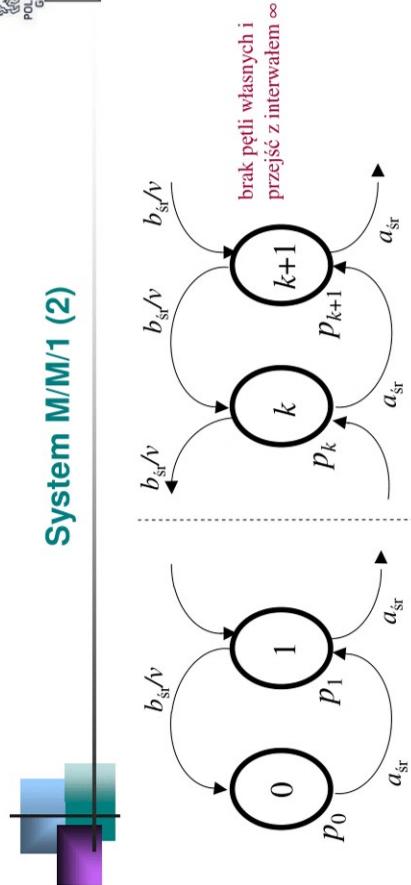
- Rozkład wykładniczy jest bezpamięciowy (własność Markowa)

⇒ interwał resztkowy (czas do wystąpienia najbliższego zdarzenia urodziny / śmierć jest statystycznie taki sam jak interwał pomiędzy takimi zdarzeniami

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

86

## System M/M/1 (2)



Sugestynna interpretacja "hydrauliczna" :

- "masa prawdopodobieństwa" ~ ciecz
- stan systemu ~ zbiornik cieczy
- $p_k$  ~ ciśnienie cieczy w zbiorniku
- przejście stanów = rura
- średnią interwał resztkowy (czas do wystąpienia) zdarzenia wyzwalającego przejęście stanów  
~ opór przepływu (=1/przekrój rury)

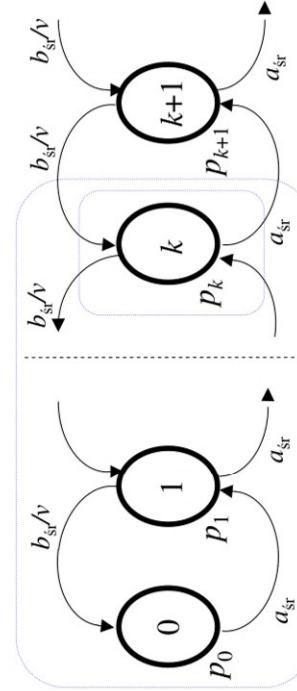
J. Komorski, Badania operacyjne/Systemy masowej obsługi I

87

## System M/M/1 (3)

W równowadze statystycznej obowiązuje bilans przepływu cieczy przez dowolny kontur zamknięty. (To są właśnie nasze równania urodzin i śmierci.)

Wybieramy kontury przecinające jak najmniej strzałek!



$$\frac{p_k}{a_{sr}} = \frac{p_{k+1}}{(b_{sr}/v)} \Rightarrow \frac{p_{k+1}}{p_k} = \frac{b_{sr}}{a_{sr}v} = r$$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

88

### System M/M/1 (4)

$$p_k = p_0 \cdot r^k \quad p_0 + p_1 + p_2 + \dots = 1 \Rightarrow p_0 = 1 - r$$

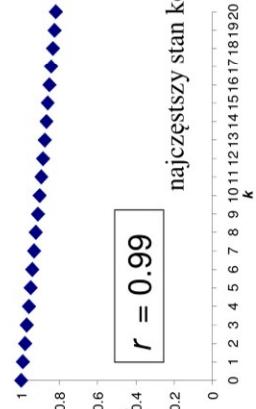
Stąd średnia długość kolejki i z tzw. Little'a średnie opóźnienie buforowania:

$$N_{\text{sr}} = \frac{r}{1-r}$$

$$d_{\text{sr}} = a_{\text{sr}} N_{\text{sr}} = \tau_{\text{sr}} \cdot \frac{1}{1-r}$$

$$w_{\text{sr}} = d_{\text{sr}} - \tau_{\text{sr}}$$

*b. sugestywne!*

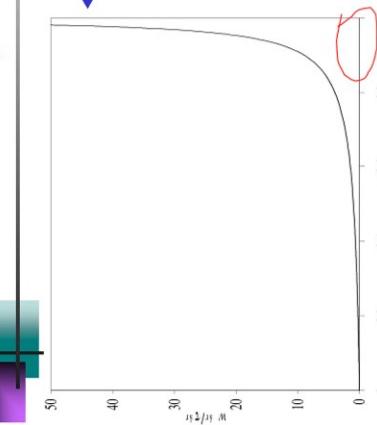


89

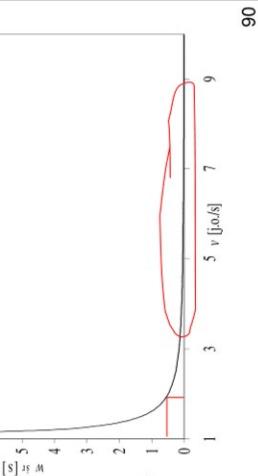
J. Komorski, Badania operacyjne/Systemy masowej obsługi I

### System M/M/1 (5)

$w_{\text{sr}}/\tau_{\text{sr}}$  jako funkcja  $r$ :  
niebezpieczny zakres obciążen



$w_{\text{sr}}$  jako funkcja  $v$ :  
nieopłacalność inwestycji w procesor



J. Komorski, Badania operacyjne/Systemy masowej obsługi I

90

## Systemy M/M/...

Okazuje się, że proces urodzin i śmierci nadaje się do analizy o wiele bogatszych (i bardziej realistycznych) markowowskich modeli systemów masowej obsługi, np.:

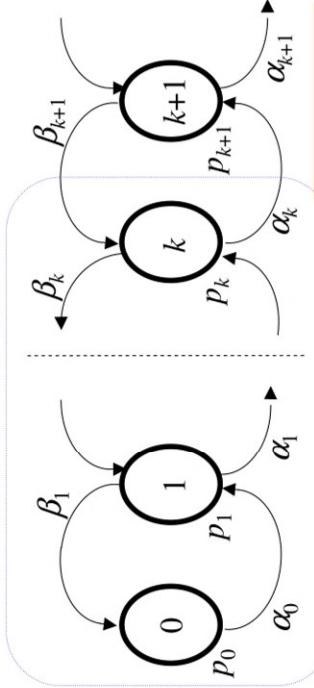
- skończona pojemność pamięci buforowej (**bez ponawiania, drop-tail**)
- wiele procesorów (wymagania: dowolny procesor), być może w liczbie zależnej od stanu kolejki
- intensywność przybywania zgłoszeń zależy od stanu kolejki (źródła zgłoszeń modelowane jako inteligentne końcówki)
- różne specjalne modele – postój taksówek / wiadro tokenowe, niesierpliwość, ...  
...i to praktycznie *bez komplikacji analizy!*

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

91

## Systemy M/M/... (2)

Średni interwał przejścia może zależeć od stanu wyjściowego:



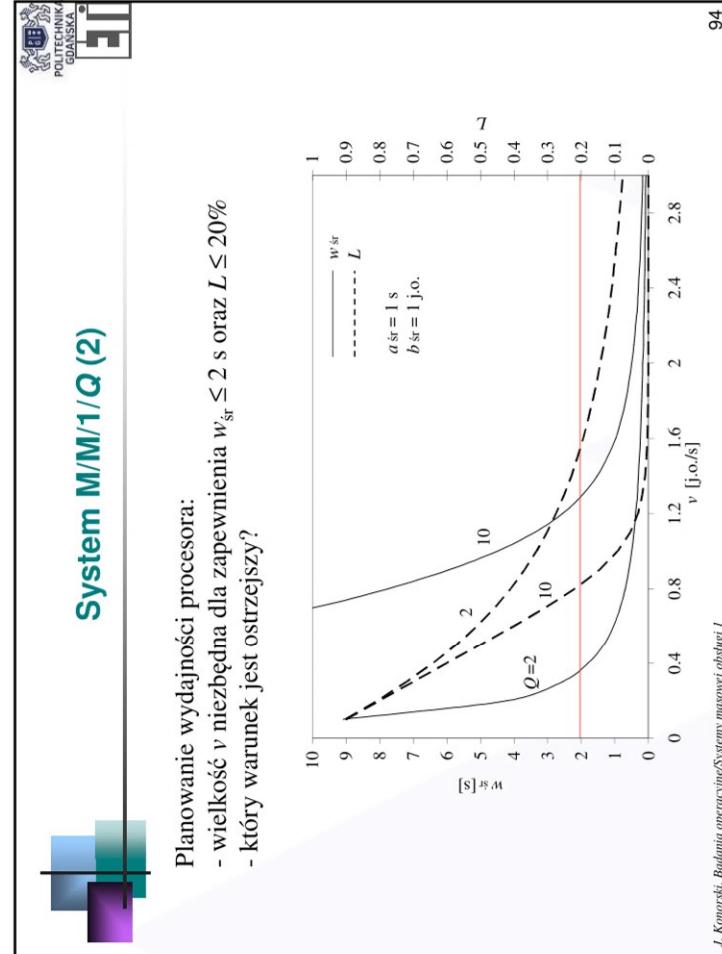
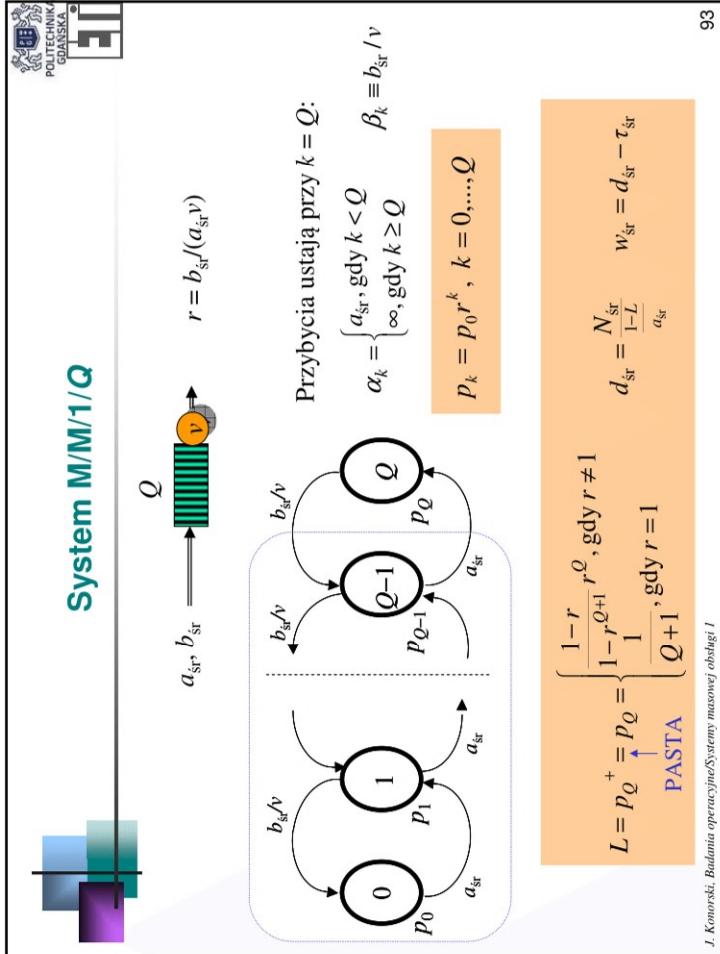
$$p_k/\alpha_k = p_{k+1}/\beta_{k+1}$$

Rozwiązań:  $p_k = p_0 \frac{\beta_1 \dots \beta_k}{\alpha_0 \dots \alpha_{k-1}}, \quad k = 0, 1, 2, \dots \quad p_0$  z warunku normalizacyjnego

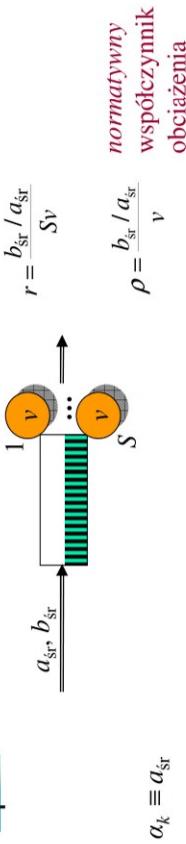
Można stąd wyznaczyć wszystkie interesujące charakterystyki:  $L, p_0, N_{\text{sr}}, d_{\text{sr}}, W_{\text{sr}}, \dots$

92

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

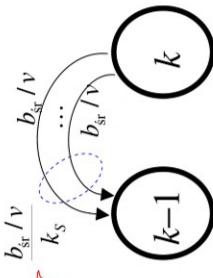


## System M/M/S



Spośród  $k$  zgłoszeń w systemie w trakcie obsługi jest  $\min\{k, S\} = k_S \Rightarrow \beta_k = \frac{b_{sr} / v}{k_S}$

$k_S$  przejść o średnich  
intervallach rezkowanych  
(czasach do wystąpienia) =  $b_{sr} / v$

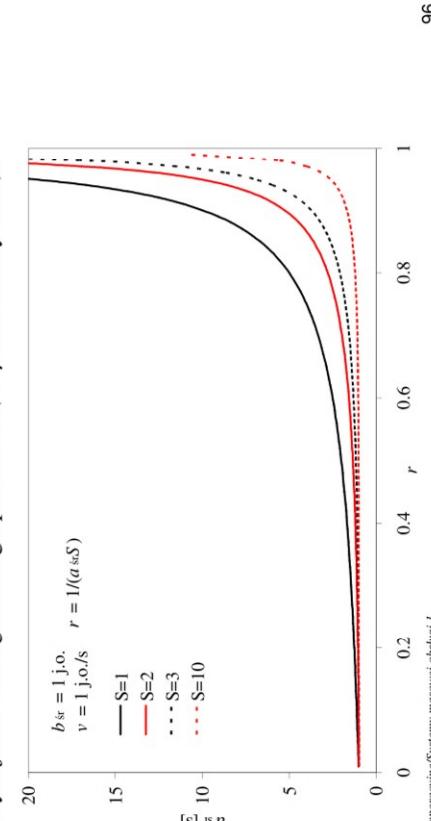


J. Komorski, Badania operacyjne/Systemy masowej obsługi I 95

## System M/M/S (2)

$$p_k = p_0 \frac{\rho^k}{k_S! S^{k-k_S}} \quad \begin{matrix} \text{Stad } N_{sr} \\ \text{Dalej z prawa Little'a, } d_{sr}. \end{matrix}$$

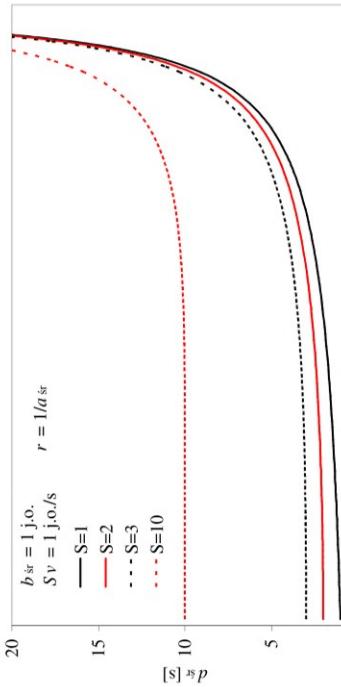
Stała wydajność obsługi każdego procesora ( $\Rightarrow \rho$  niezależny od  $S$ ):



J. Komorski, Badania operacyjne/Systemy masowej obsługi I 96

### System M/M/S (3)

Stała łączna wydajność obsługi ( $\Rightarrow \rho$  rośnie wraz z S):

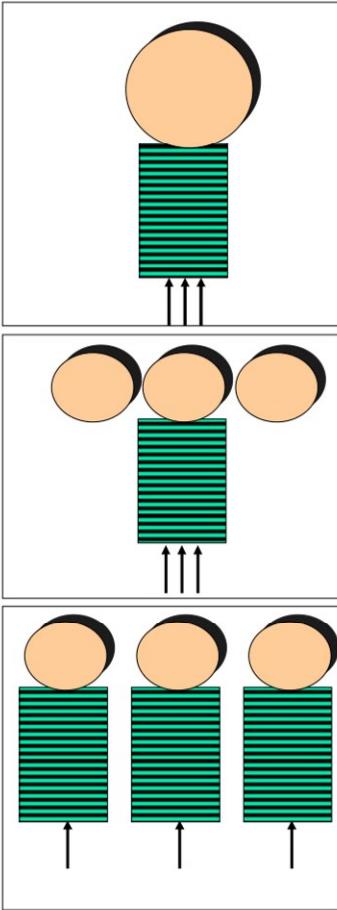


J. Komorski, Badania operacyjne/ Systemy masowej obsługi I 97

(Widzieliśmy już poprzednio, że rozdrobnienie podaży obsługi szkodzi!)

### O pozytku ze współdzielenia bufora i komasacji podaży obsługi

$$r = 0.667$$



$3 \times M/M/1$

$$d_{sr} = 1 \text{ s}$$

$M/M/3$

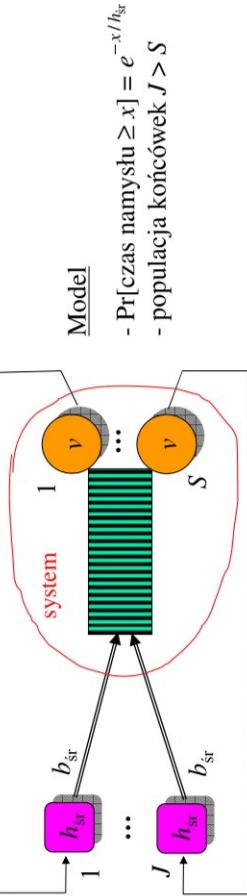
$$d_{sr} = 0.48 \text{ s}$$

$M/M/1$

$$d_{sr} = 0.33 \text{ s}$$

J. Komorski, Badania operacyjne/ Systemy masowej obsługi I 98

## System M/M/S/J



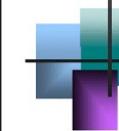
Gdy  $k$  zgłoszeń w systemie:

- w trakcie namysłu  $J - k$  końcowek  $\Rightarrow \alpha_k = \frac{h_{sr}}{J - k}$
- w trakcie obsługi  $k_s$  zgłoszeń  $\Rightarrow \beta_k$  jak dla M/M/S
- $J \rightarrow \infty, h_{sr} \rightarrow \infty, h_{sr}/J \rightarrow \alpha_{sr}, M/M/S$

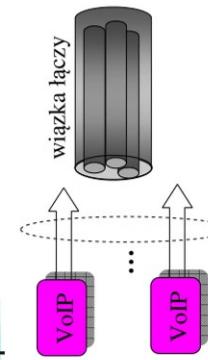
J. Komorski, Badania operacyjne i Systemy masowej obsługi I

99

## System M/M/S/S



brak oczekiwania w kolejce  
 $v$  – prędkość transmisji w 1 łączu  
 $\rho = b_{sr}/(a_{sr}v)$   
– normatywny współczynnik obciążenia  
 $\alpha_k, \beta_k, k = 0, \dots, S$  jak dla M/M/S



$$p_k = p_0 \frac{\rho^k}{k!} \quad p_S = L = \sum_{k=0}^S \frac{\rho^k}{k!}$$

• styczna formuła B Erlanga

- magia: prawdziwa dla dowolnego rozkładu wymagań, tj. dla M/G/S/S ( )
- istnieją kalkulatory online ([www.volp-calculator.com/calculator/](http://www.volp-calculator.com/calculator/))

J. Komorski, Badania operacyjne i Systemy masowej obsługi I

100

## System M/M/ $\infty$

Do bardzo dużego hipermarketu wchodzi średnio 20 klientów na minutę, a każdy średnio robi zakupy (łącznie z czekaniem przy kasie) przez 15 minut.

Oblicz rozkład liczby klientów w hipermarketie w dowolnej chwili, przyjmując markowski model systemu obsługi.

$$a_{\text{sr}} = 3 \text{ s}, \tau_{\text{sr}} = 900 \text{ s}, \rho = 300 \text{ erlangów}$$

$$N_{\text{sr}} = \rho = 300 \text{ (to wiemy z prawa Little'a)}$$

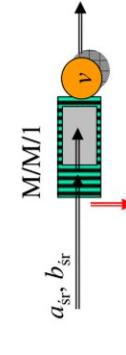
Szczególny przypadek M/M/S/S;  $S \rightarrow \infty$ , zatem  $p_0 = \frac{1}{\sum_{k=0}^{\infty} \frac{\rho^k}{k!}} = e^{-\rho}$

$$p_k = e^{-\rho} \frac{\rho^k}{k!} - \text{rozkład Poissona (!)}$$

*J. Komorski, Badania operacyjne i Systemy masowej obsługi I*

101

## Niecierpliwe zgłoszenia



prawdopodobieństwo rezygnacji  
 $= k/(k+1)$ ,  $k$  – liczba zastanych zgłoszeń

Przyjmując model markowski, oblicz frakcję ucieczek.

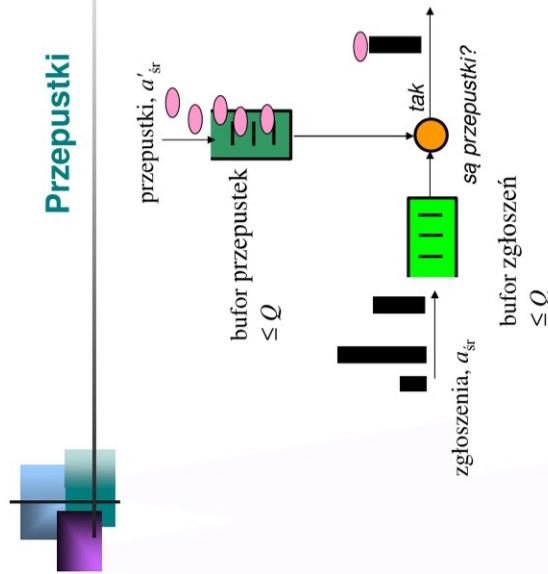
$$\alpha_k = \frac{a_{\text{sr}}}{1 - \frac{k}{k+1}} = a_{\text{sr}}(k+1), \quad \beta_k = b_{\text{sr}}/v \Rightarrow (p_k) - \text{rozkład Poissona (!)}$$

$$L = p_0 \cdot 0 + p_1 \cdot \frac{1}{2} + p_2 \cdot \frac{2}{3} + \dots + p_k \cdot \frac{k}{k+1} + \dots$$

*J. Komorski, Badania operacyjne i Systemy masowej obsługi I*

102

## Przepustki



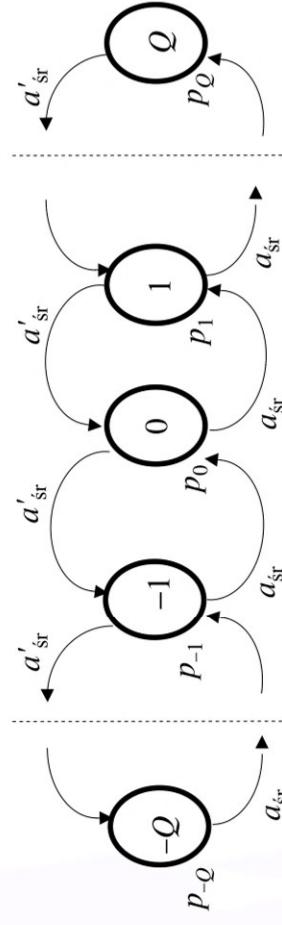
Przyjmując model markowski, oblicz prawdopodobieństwo, że w równowadze statystycznej przepustki czekają na zgłoszenia/zgłoszenia czekają na przepustki.

J. Komorski, Badania operacyjne/Syttemy masowej obsługi I

103

## Przepustki (2)

Nowe zgłoszenie zwiększa stan systemu, nowa przepustka zmniejsza:

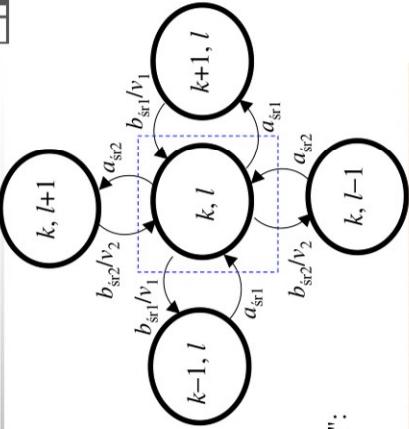
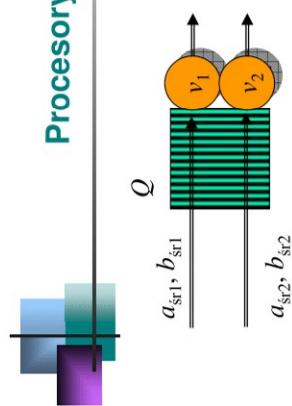


Czy przy  $Q = \infty$  wystąpi równowaga statystyczna?

J. Komorski, Badania operacyjne/Syttemy masowej obsługi I

104

## Procesory dedykowane



Bilans przepływu "masy prawdopodobieństwa":

$$p_{k,l}/a_{sr1} + p_{k,l}/a_{sr2} + p_{k,l}/(b_{sr1}/v_1) + p_{k,l}/(b_{sr2}/v_2) = p_{k-1,l}/a_{sr1} + p_{k,l-1}/a_{sr2} + p_{k+1,l}/(b_{sr1}/v_1) + p_{k,l+1}/(b_{sr2}/v_2)$$

$$\text{Rozwiążanie przez odgadnięcie: } p_{k,l} = p_{0,0} r_1^k r_2^l, \text{ gdzie } r_1 = \frac{b_{sr1}}{a_{sr1} v_1}, r_2 = \frac{b_{sr2}}{a_{sr2} v_2}$$

J. Komorski, Badania operacyjne/Sytetem masowej obslugi I

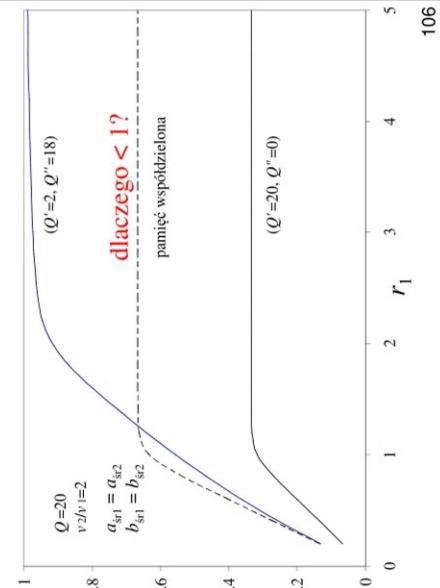
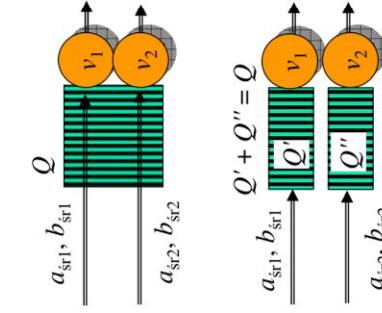
105

## O pozytku z rozdziału pamięci buforowej

$$\frac{1}{b_{sr1}/v_1} \left( 1 - \sum_{l=0}^{\varrho} p_{0,l} \right) + \frac{1}{b_{sr2}/v_2} \left( 1 - \sum_{k=0}^{\varrho} p_{k,0} \right)$$

$$\frac{1}{b_{sr1}/v_1} + \frac{1}{b_{sr2}/v_2}$$

Znormalizowana przepływność:



J. Komorski, Badania operacyjne/Sytetem masowej obslugi I

106

## Transformata Laplace'a

Wyznaczanie rozkładu prawdopodobieństwa opóźnien jest trudniejszym zadaniem, wymaga dodatkowego aparatu matematycznego – transformata Laplace'a.

Dla zmiennnej losowej  $X$  o komplementarnej dystrybuancie  $P(x)$  definiujemy:

$$X^*(s) = \int_0^\infty e^{-sx} (-dP(x)) = \left( e^{-sx} \right)_{\text{sr}}$$

Np. dla rozkładu wykładniczego  $P(X \geq x) = e^{-x/c} \Rightarrow X^*(s) = \frac{1}{cs+1}$

Transformata Laplace'a jest operatorem liniowym.

Jeżeli zmienne losowe  $X_1$  i  $X_2$  są statystycznie niezależne oraz  $X = X_1 + X_2$ , to

$$X^*(s) = X_1^*(s)X_2^*(s).$$

Dla sumy  $2, 3, \dots$  zmiennych losowych o rozkłach wykładniczych :

$$X^*(s) = \left( \frac{1}{cs+1} \right)^2, \left( \frac{1}{cs+1} \right)^3, \dots$$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

107

## Transformata Laplace'a (2)

Mając  $X^*(s)$ , jak znaleźć  $P(x)$ ? Czyli jak odwzorować transformatę Laplace'a?

- Można analitycznie – całka Bromwicha... niezalecane :)

- Pomocne są obszerne tablice, np. gdy  $X^*(s) = \left( \frac{1}{cs+1} \right)^K$  ( $K$  całkowite), to

$$P(x) = e^{-x/c} \sum_{i=0}^{K-1} \frac{(x/c)^i}{i!} \quad (x \geq 0) \text{ rozkład Erlanga rzędu } K$$

- Istnieje też dużo software'u rachunku symbolicznego, por. np. [www.educypedia.be/education/calculator/salgebra.htm](http://www.educypedia.be/education/calculator/salgebra.htm)

- Dla pewnych typów  $X^*(s)$  powyższe zawodzi. Zostają algorytmy numeryczne. Nie ma jednego uniwersalnego, bo problem jest numerycznie niestabilny! Bezpośrednie korzystanie ze wzoru na transformatę odwrotną jest kłopotliwe, gdyż wymaga operowania na liczbach zespolonych.

J. Komorski, Badania operacyjne/Systemy masowej obsługi I

108



## Transformata Laplace'a (3)

[www.pe.tamu.edu/bblasingame/data/P620\\_reference/data/P620\\_LectureNotes\\_\(pdf\)/P620\\_Mod1\\_Math/P620\\_ML\\_05\\_LaplaceTrans.pdf](http://www.pe.tamu.edu/bblasingame/data/P620_reference/data/P620_LectureNotes_(pdf)/P620_Mod1_Math/P620_ML_05_LaplaceTrans.pdf)

- The Gaver formula for numerical Laplace transform inversion is

$$f_{Gaver}(n, t) = \frac{\ln(2)}{t} \frac{(2n)!}{(n-1)!} \sum_{k=0}^n \frac{(-1)^k}{(n-k)!k!} f\left[\frac{\ln(2)}{t} (n+k)\right]$$

- The Gaver-Stehfest formula for numerical Laplace transform inversion is

$$f_{Gaver-Stehfest}(n, t) = \frac{\ln(2)}{t} \sum_{i=1}^n V_i f\left[\frac{\ln(2)}{t} i\right]$$

and the Stehfest extrapolation coefficients are given

$$V_i = (-1)^{\frac{n}{2}+i} \sum_{k=\lceil \frac{i+1}{2} \rceil}^{\frac{n}{2}} \frac{\frac{n}{2} (2k)!}{k! (k-1)!(i-k)!(2k-i)!}$$

J. Komorski, Badania operacyjne/Systemy masowej obsługi I 109



## M/M/1 FIFO: rozkład opóźnień systemowych

Opóźnienie systemowe zgłoszenia zastającego  $k$  zgłoszeń w chwili przybycia składa się z  $k + 1$  czasów obsługi (w tym jednego resztowego, gdy  $k > 0$ ) o rozkładach wykładniczych  $P(x) = e^{-x/\tau_m}$ .

Zatem jego transformata Laplace'a:  $\left(\frac{1}{\tau_{sr}s + 1}\right)^{k+1}$

Działa PASTA, zatem  $p_k^+ \equiv p_k$ .

Uśredniając na różne liczby zastanych zgłoszeń i wykorzystując liniowość:

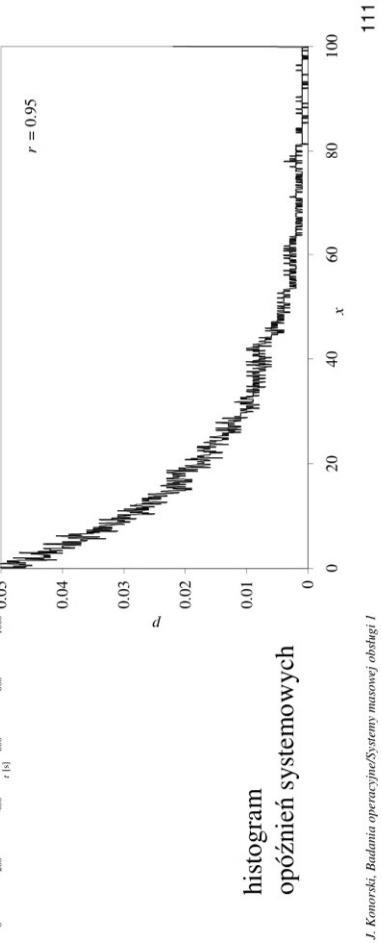
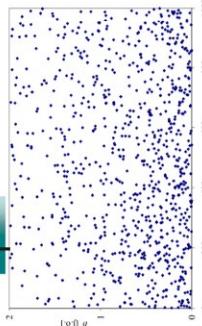
$$D^*(s) = \sum_{k=0}^{\infty} p_k \left( \frac{1}{\tau_{sr}s + 1} \right)^{k+1} = \sum_{k=0}^{\infty} (1-r)r^k \left( \frac{1}{\tau_{sr}s + 1} \right)^{k+1} = \frac{1}{\frac{\tau_{sr}}{1-r}s + 1}$$

Rozkład wykładniczy (o znanej już wcześniej wartości średniej) !

J. Komorski, Badania operacyjne/Systemy masowej obsługi I 110

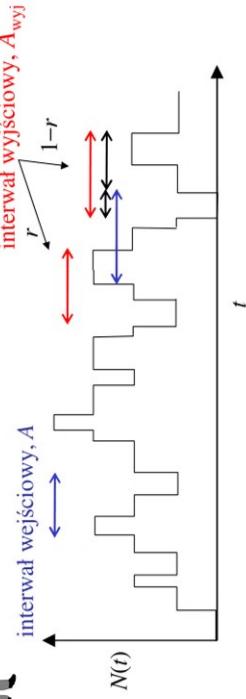
## M/M/1 FIFO: rozkład opóźnień systemowych (2)

strumień M/M,  $a_{\text{sr}} = b_{\text{sr}} = 1$



## M/M/1: strumień wyjściowy

Jak wygląda strumień wyjściowy z M/M/1?



Działa PASTA i argument skoków  $\pm 1$ , zatem  $p_k \equiv p_k^+ \equiv p_k^-$ . Czyli  $p_0^- = 1 - r$ .

$A_{\text{wyj}}$  jest z prawdopodobieństwem  $r$  jednym czasem obsługi, zaś z prawd.  $1 - r$  sumą jednego czasu obsługi i resztkowego interwala wejściowego.

Wyznaczenie rozkładu  $A_{\text{wyj}}$  jest nietrudne, gdy już się opanowało transformatę Laplace'a.

## M/M/1: strumień wyjściowy (2)

$$A_{\text{wyj}}^*(s) = r \frac{1}{\tau_{\text{st}} s + 1} + (1 - r) \frac{1}{a_{\text{st}} s + 1} \cdot \frac{1}{\tau_{\text{st}} s + 1} = \frac{1}{a_{\text{st}} s + 1} = A^*(s)$$

