Στοιχεία Θεωρίας Αναμονής (queueing theory)

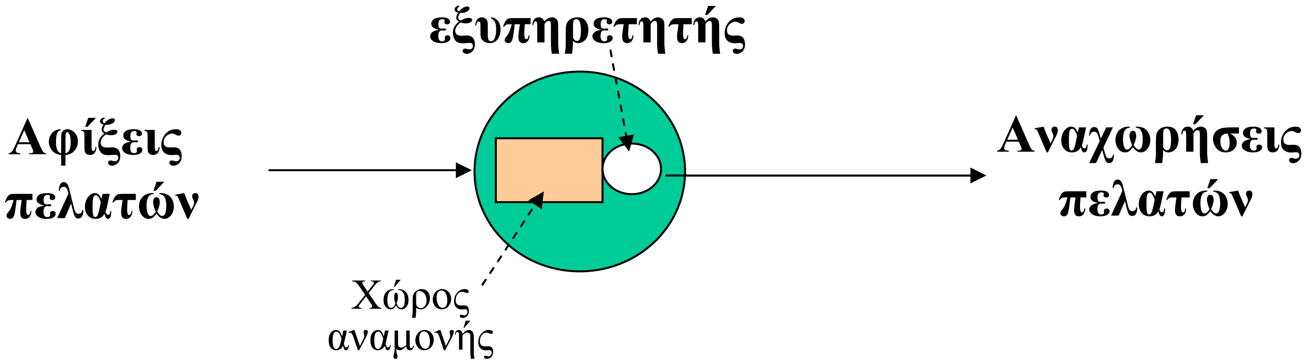
Θεωρία Αναμονής

- *Βασικό μαθηματικό εργαλείο για την ανάλυση της επίδοσης και το σχεδιασμό δικτύων,*

  αφού η ζήτηση (κοινόχρηστων) δικτυακών υπηρεσιών και πόρων είναι διαδικασία στοχαστική και δημιουργεί ουρές αναμονής

- *Η θεμελίωση και ανάπτυξη της θεωρίας αναμονής οφείλεται εν πολλοίς στις ανάγκες ανάλυσης και σχεδιασμού τηλεπικοινωνιακών δικτύων*

Αναμονητικά συστήματα

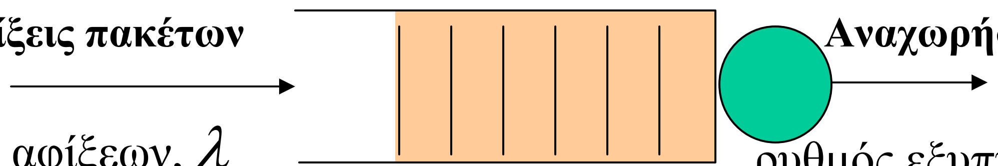


<u>Παραδείγματα</u>

1. **Γκισέ ταχυδρομείου**
2. **Κόμβος μεταγωγής τηλεφωνικών κλήσεων**
3. **Κόμβος μεταγωγής πακέτων στο Διαδίκτυο**

Ένα απλό αναμονητικό σύστημα: γραμμή αποστολής πακέτων

Αφίξεις πακέτων **Αναχωρήσεις πακέτων**



ρυθμός αφίξεων, λ

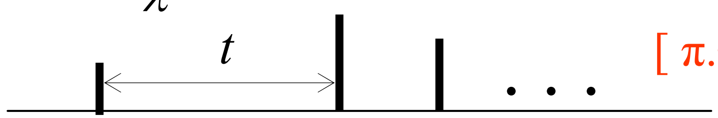
(arrival rate, packets/sec)

ρυθμός εξυπηρέτησης, μ

(service rate, packets/sec)

arrival process

$\frac{1}{\lambda}$: mean interarrival distance



service-time distribution

$\frac{1}{\mu}$: mean service time

[π.χ. πακέτα των 2000 bits σε γραμμή 2400 bps

==> μ = 1.2 packets/sec]

Χρησιμοποίηση γραμμής:

(link utilization or traffic intensity)

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

Για ευσταθές σύστημα πρέπει: ρ<1, ή

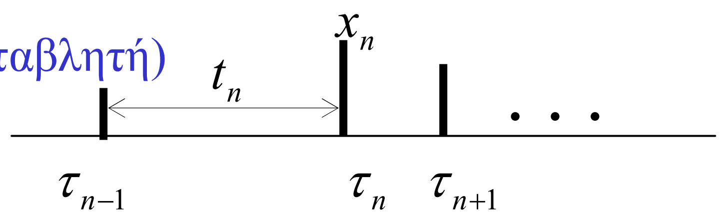
πεπερασμένος χώρος αναμονής & φραγή

Παράμετροι απλού αναμονητικού συστήματος

- Παράμετροι διαδικασίας αφίξεων πελατών
  - Κατανομή ενδοδιαστημάτων αφίξεων
  - Συσχέτιση »
- Παράμετροι χρόνου εξυπηρέτησης πελατών
- Τρόπος εξυπηρέτησης (service discipline)
  - FIFO (First In First Out)
  - LIFO (Last In First Out)
  - Προτεραιότητες
  - ...

Συμβολισμοί

- $\tau_n =$ χρόνος άφιξης n^οστού^ πελάτη (C$_n$: πακέτο, κλήση κλπ)
- $t_n = \tau_n - \tau_{n-1} =$ ενδοδιάστημα μεταξύ αφίξεων πελατών C_n και C_{n-1}
- $x_n =$ χρόνος εξυπηρέτησης πελάτη C_n
- $\tilde{t} =$ ενδοδιάστημα μεταξύ αφίξεων (τ. μεταβλητή)
- $\tilde{x} =$ χρόνος εξυπηρέτησης (τ. μεταβλητή)



- $A(t) = \Pr\{\tilde{t} \le t\}$ αθροιστική συνάρτηση κατανομής πιθ. ενδοδιαστημάτων αφ.
- $B(x) = \Pr\{\tilde{x} \le x\}$ αθροιστική συνάρτηση κατανομής πιθ. χρόνου εξυπηρέτησης
- $a(t) = dA(t)/dt$ συνάρτηση πυκνότητας πιθανότητας ενδοδιαστημάτων αφ.
- $b(x) = dB(x)/dx$ συνάρτηση πυκνότητας πιθανότητας χρόνου εξυπηρέτησης

- $E\{\tilde{t}\} = \bar{t} = 1/\lambda$ $\qquad E\{\tilde{x}\} = \bar{x} = 1/\mu$

- $N(t) =$ αριθμός πελατών στο σύστημα τη χρονική στιγμή t
- $w_n =$ χρόνος αναμονής (στην ουρά) του C_n $\quad s_n =$ χρόνος στο σύστημα (αναμονής + εξυπηρέτησης)

Συμβολισμοί (II)

Kendal's notation:

A / B / m

κατανομή ενδοδιαστημάτων

αριθμός εξυπηρετητών

κατανομή χρόνου εξυπηρ.

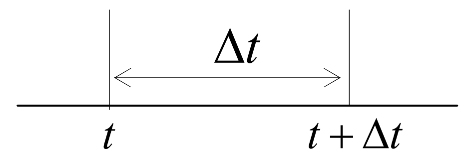
όπου A,B ένα από:

M = exponential (i.e. **M**arkovian)
E_r=r-stage **E**rlangian
H_r=r-stage **H**yperexponential
D =**D**eterministic
G =**G**eneral

π.χ. M/M/1: αφίξεις Poisson, εκθετικοί χρόνοι εξυπ., 1 εξυπηρετητής

Διαδικασία αφίξεων Poisson

- $\Pr\{\text{one arrival} \in \Delta t\} = \lambda\Delta t + 0(\Delta t)$
- $\Pr\{\text{no arrivals} \in \Delta t\} = 1 - \lambda\Delta t + 0(\Delta t)$
- Arrivals are memoryless



$$\Pr\{k \ arrivals \in T\} = \frac{(\lambda T)^k e^{-\lambda T}}{k!}$$

(Poisson distribution)

$$E\{k\} = \sum_{k=0}^{\infty} kP_k = \lambda T \qquad \sigma_k^2 \equiv E\{k - E\{k\}\}^2 = \lambda T$$

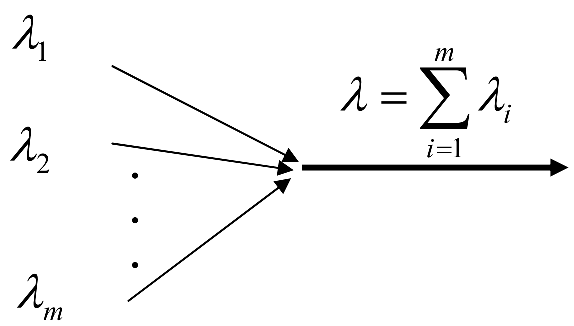
$$a(t) = \lambda e^{-\lambda t}$$

(exponential interarrival time pdf)

$$E\{\tilde{t}\} = \int_0^{\infty} t a(t) dt = \frac{1}{\lambda} \qquad \sigma_{\tilde{t}}^2 \equiv E\{\tilde{t} - E\{\tilde{t}\}\}^2 = \frac{1}{\lambda^2}$$

Διαδικασία αφίξεων Poisson (II)

- Η υπέρθεση ροών Poisson είναι επίσης Poisson με ρυθμό λ ίσο με το άθροισμα των ρυθμών των επιμέρους ροών



Γενικές σχέσεις (G/G/1)

$$\rho = \lambda \bar{x} = \lambda / \mu$$

χρησιμοποίηση (utilization)

$$0 \le \rho < 1$$

για ευσταθές σύστημα

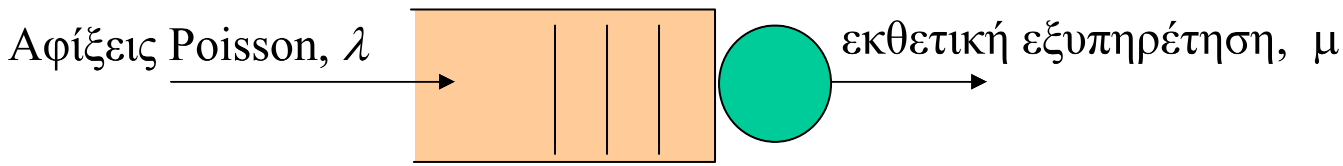
$$T = \bar{x} + W$$

μέσος χρόνος παραμονής στο σύστημα
(average time in the system)

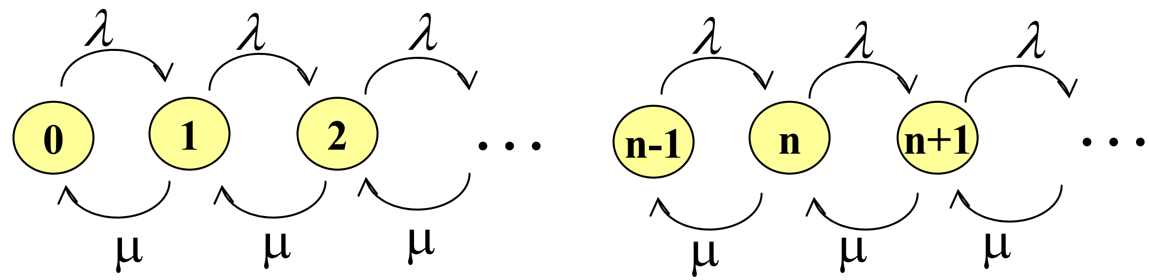
$$\bar{N} = \lambda T$$

μέσος αριθμός πελατών στο σύστημα
(average number of customers in the system)
(Little's formula)

M/M/1 queue



state diagram



$$p_n(t+\Delta t) = p_n(t)[(1-\lambda\Delta t)(1-\mu\Delta t) + \lambda\Delta t\mu\Delta t + 0(\Delta t)]$$
$$+\, p_{n-1}(t)[\lambda\Delta t(1-\mu\Delta t) + 0(\Delta t)]$$
$$+\, p_{n+1}(t)[\mu\Delta t(1-\lambda\Delta t) + 0(\Delta t)]$$

(equilibrium)*

balance equation $(\lambda+\mu)p_n = \lambda p_{n-1} + \mu p_{n+1}\ , \quad n \ge 1$

(*) κατάσταση ισορροπίας: $[p_n(t+\Delta t) - p_n(t)]/\Delta t \xrightarrow[\Delta t \to 0]{} 0$

M/M/1 queue (II)

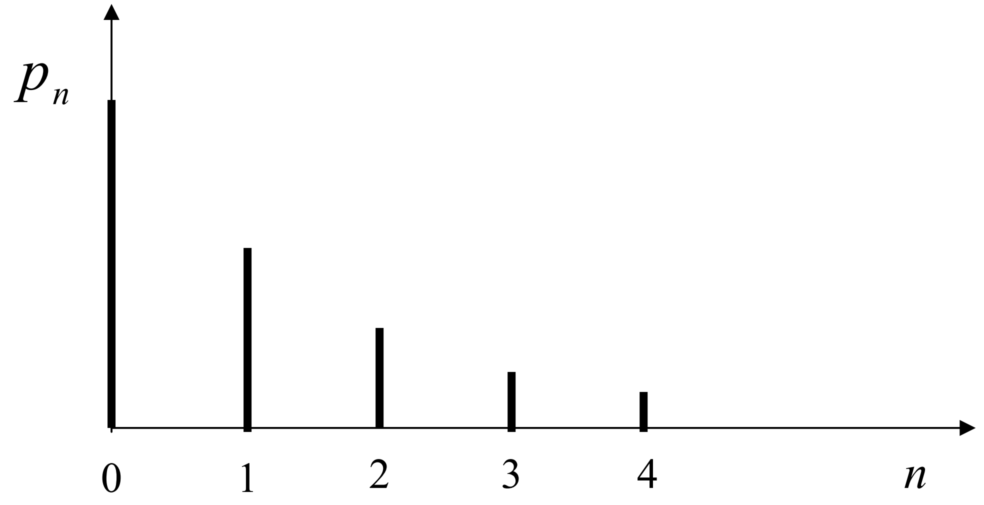
Λύση (γεωμετρική κατανομή)

$$\begin{cases} p_n = p_o \rho^n, \quad \rho = \lambda / \mu \\ \sum_n p_n = 1 \end{cases} \Rightarrow \quad p_n = (1-\rho)\rho^n, \quad \rho = \lambda/\mu < 1$$

$$\overline{N} = \frac{\rho}{1-\rho}$$

$$W = \frac{\rho/\mu}{1-\rho}$$

$$T = W + 1/\mu = \frac{1/\mu}{1-\rho}$$



Παράδειγμα M/M/1

ταχύτητα γραμμής = 2400 bps
μέσο μήκος πακέτων = 1000 bits

$$\left.\begin{array}{l} \text{ταχύτητα γραμμής} = 2400 \text{ bps} \\ \text{μέσο μήκος πακέτων} = 1000 \text{ bits} \end{array}\right\} \mu = 2.4 \text{ packets/sec}$$

a) $\underline{\lambda = 1 \text{ packet/sec}}$

$$\rho = 1/2.4 = 0.42, \quad \overline{N} = \rho/(1-\rho) = 0.72, \quad T = \overline{N}/\lambda = 0.72 \text{ sec}$$

b) $\underline{\lambda = 2 \text{ packets/sec}}$

$$\rho = 2/2.4 = 0.833, \quad \overline{N} = \rho/(1-\rho) = 5, \quad T = \overline{N}/\lambda = 2.5 \text{ sec}$$

c) $\underline{\lambda = 2.3 \text{ packets/sec}}$

$$\rho = 2.3/2.4 = 0.958, \quad \overline{N} = \rho/(1-\rho) = 22.98, \quad T = \overline{N}/\lambda = 10 \text{ sec}$$

M/M/1/N queue

(M/M/1 queue with a finite buffer space – N places)

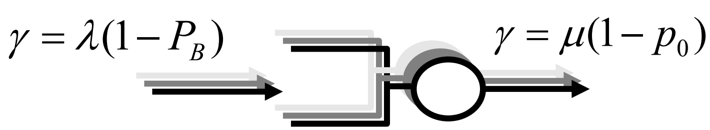
$$
\begin{cases}
p_n = p_o \rho^n, \quad \rho = \lambda / \mu \\
\sum_{n=1}^{N} p_n = 1
\end{cases}
\Rightarrow
\boxed{p_n = \frac{(1-\rho)\rho^n}{1-\rho^{N+1}}, \quad \rho = \lambda / \mu < 1}
$$

Πιθανότητα φραγής (blocking probability):

$$P_B = p_N = \frac{(1-\rho)\rho^N}{1-\rho^{N-1}} \underset{\rho^{N+1} << 1}{\approx} (1-\rho)\rho^N$$

Διέλευση (throughput):

$$\gamma = \lambda(1-P_B) = \mu(1-p_0)$$

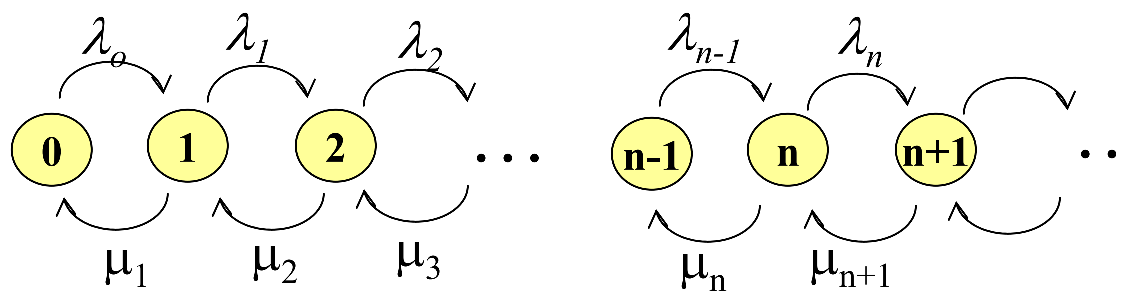


$$\boxed{\frac{\gamma}{\mu} = (1-p_o) = \frac{\rho(1-\rho^N)}{1-\rho^{N+1}}}$$

Birth-Death processes

State-dependent queues

State diagram



Balance equation

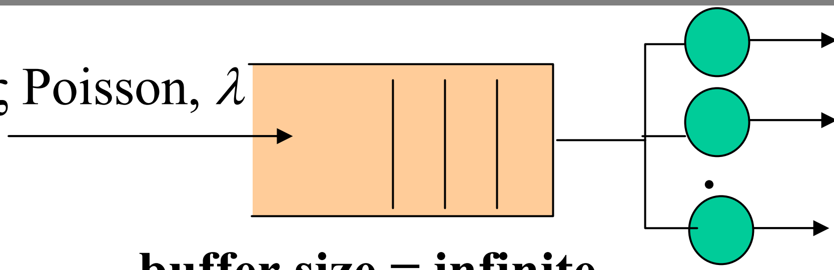
$$(\lambda_n + \mu_n) p_n = \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1} \,, \quad n \geq 1$$

$$\lambda_n p_n = \mu_{n+1} p_{n+1}$$

Queue size = *N*

$$p_n / p_0 = \prod_{i=0}^{n-1} \lambda_i \Big/ \prod_{i=1}^{n} \mu_i \,, \qquad \sum_{n=0}^{N} p_n = 1$$

The M/M/m queueing system



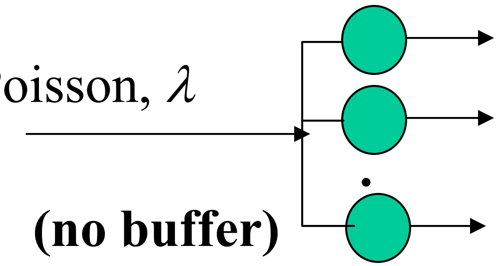
(Erlang-C formula)

$$p_k = \begin{cases} p_0 \dfrac{(m\rho)^k}{k!}, & k \le m \\ p_0 \dfrac{\rho^k m^m}{m!}, & k \ge m \end{cases} \qquad with \qquad p_0 = \left[\sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!(1-\rho)} \right]^{-1}$$

όπου: $\rho = \dfrac{\lambda}{m\mu} < 1$

The M/M/m/m queueing system

αφίξεις Poisson, λ



(no buffer)

m εκθετικοί εξυπηρ., μ

$$p_k = \frac{(\lambda/\mu)^k / k!}{\sum_{i=0}^{m} (\lambda/\mu)^i / i!}, \quad 0 \le k \le m$$

(Erlang-B or Erlang-loss formula)

πιθανότητα φραγής: $$P_B = \frac{(\lambda/\mu)^m / m!}{\sum_{i=0}^{m} (\lambda/\mu)^i / i!}$$

Παραδείγματα

Παράδειγμα 1: πιθανότητα φραγής κλήσης

Τηλεφωνικό κέντρο διαθέτει 8 γραμμές εξόδου προς ένα συγκεκριμένο προορισμό. Να βρεθεί η πιθανότητα φραγής κλήσης για ρυθμό αφίξεων λ= 1 κλήση/λεπτό προς το συγκεκριμένο προορισμό και μέση διάρκεια κλήσης 3 λεπτά.

$$P_B = \frac{(\lambda / \mu)^m / m!}{\sum_{i=0}^{m} (\lambda / \mu)^i / i!} \;, \quad \lambda / \mu = 3, \quad m = 8$$

$$\Rightarrow \quad P_B \approx 0.008$$

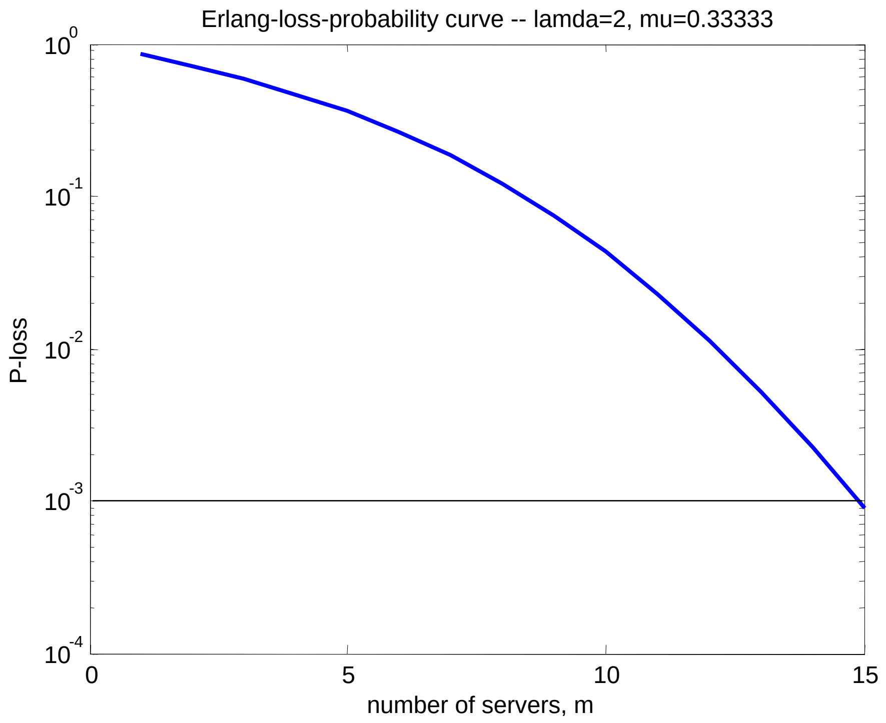
Παραδείγματα (συνέχεια)

Παράδειγμα 2: διαστασιοδότηση

Σε τηλεφωνικό κέντρο να βρεθεί ο αριθμός των γραμμών εξόδου έτσι ώστε η πιθανότητα φραγής κλήσης να είναι μικρότερη του 10^{-3} για ρυθμό αφίξεων κλήσεων $\lambda = 2$ κλήσεις/λεπτό και μέση διάρκεια κλήσης 3 λεπτά.

Παραδείγματα (συνέχεια)

```
function erlang_loss_plot(lamda,mu,m)
% m is now a vector of number of servers
% in the M/M/m/m system
m=[]; p=[];
for i=1:15
m=[m i];
p=[p erlang_loss(2,1/3,i)];
end
semilogy(m,p);
title(['Erlang-loss-probability curve --
        lamda=',num2str(lamda), ',
        mu=',num2str(mu)]);
xlabel('number of servers, m');
ylabel('P-loss');
end
```



erlang_loss_plot(2,1/3,[1:15])

Το αναμονητικό σύστημα M/G/1

Pollaczek-Khinchine formulas:

$$E\{k\} = \frac{\rho}{1-\rho}[1-\frac{\rho}{2}(1-\mu^2\sigma^2)]$$

μέσος αριθμός πελατών στο σύστημα

$$E\{T\} = \frac{E\{k\}}{\lambda} = \frac{1/\mu}{1-\rho}[1-\frac{\rho}{2}(1-\mu^2\sigma^2)]$$

μέσος χρόνος διέλευσης

λ: (μέσος) ρυθμός αφίξεων, μ: (μέσος) ρυθμός εξυπηρέτησης
$\rho=\lambda/\mu$, σ : τυπική απόκλιση χρόνου εξυπηρέτησης

Το αναμονητικό σύστημα M/G/1 (συν.)

$v = \tilde{x}C = \mu \acute{\eta} \kappa o \varsigma \ \pi \alpha \kappa \acute{\varepsilon} \tau o \upsilon$ $\qquad V^*(s) \equiv E\{e^{-sv}\}$

$W_c(x) \equiv \Pr\{\mu \acute{\eta} \kappa o \varsigma\_o \upsilon \rho \acute{\alpha} \varsigma > x\}$: η CPDF του μήκους ουράς

Τότε

$$\gamma e^{q_o x} \leq W_c(x) \leq e^{q_o x}$$

$$q_o = \sup\{q : q = \frac{\lambda}{C}(1 - V(q))\}$$