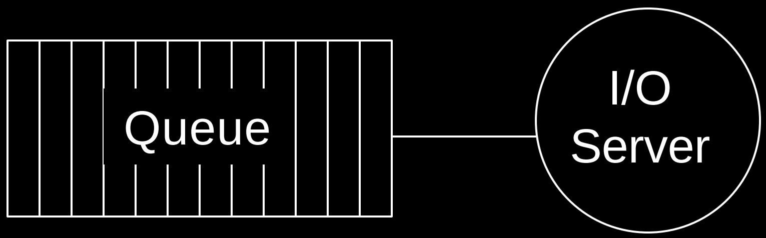
I/O Devices and Performance Measures

Lorn Miller
Steven Giovenco

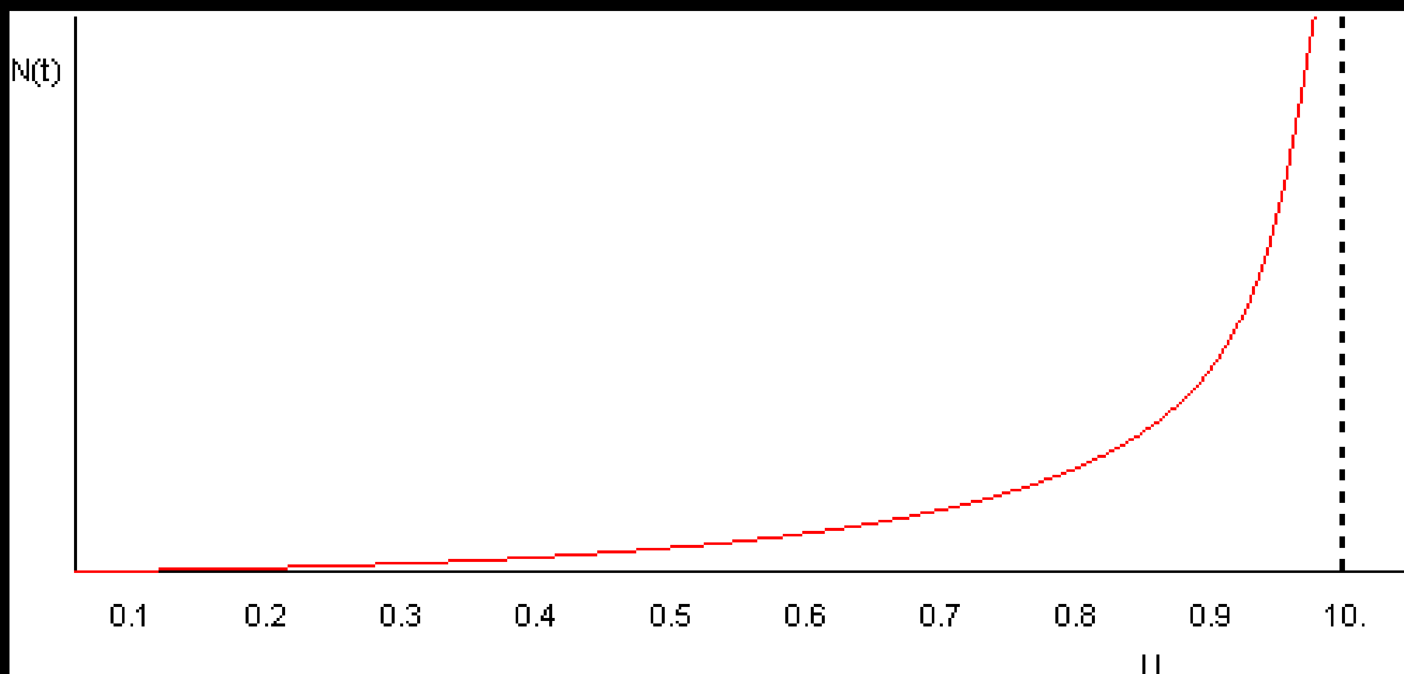
I/O Request Blocks and I/O Utilization



IORB are put in the Queue and served by the I/O Server one at a time

$\bar{N}$ is the average number of IO Request Blocks
U is the I/O device utilization (0<U<1)

$$\bar{N} = \frac{U}{1-U}$$



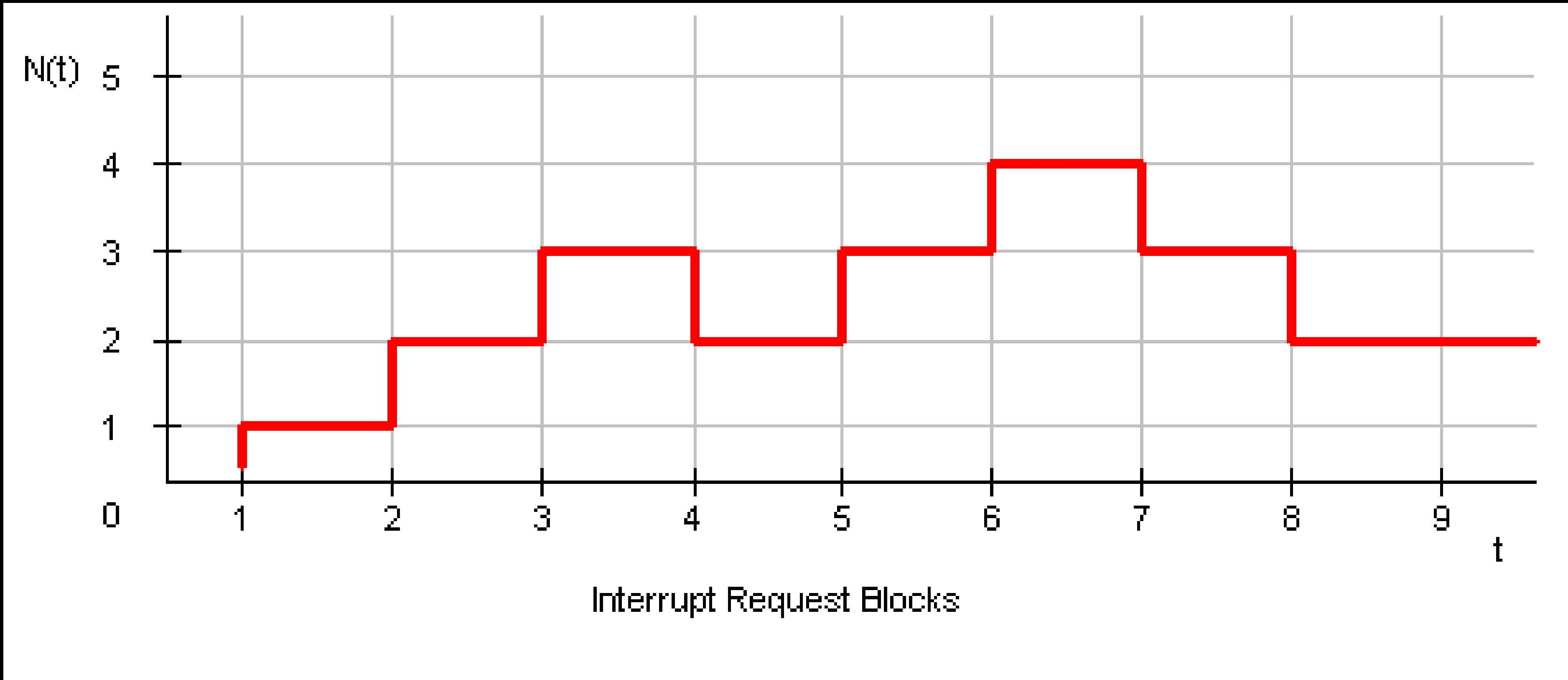
I/O Request Blocks and I/O Utilization

Example:

T0	N(0)=0	
T1	N(1)=1	one arrival
T2	N(2)=2	one arrival
T3	N(3)=3	one arrival
T4	N(4)=2	one departure
T5	N(5)=3	one arrival
T6	N(6)=4	one arrival
T7	N(7)=3	one departure
T8	N(8)=2	one departure

I/O Request Blocks and I/O Utilization

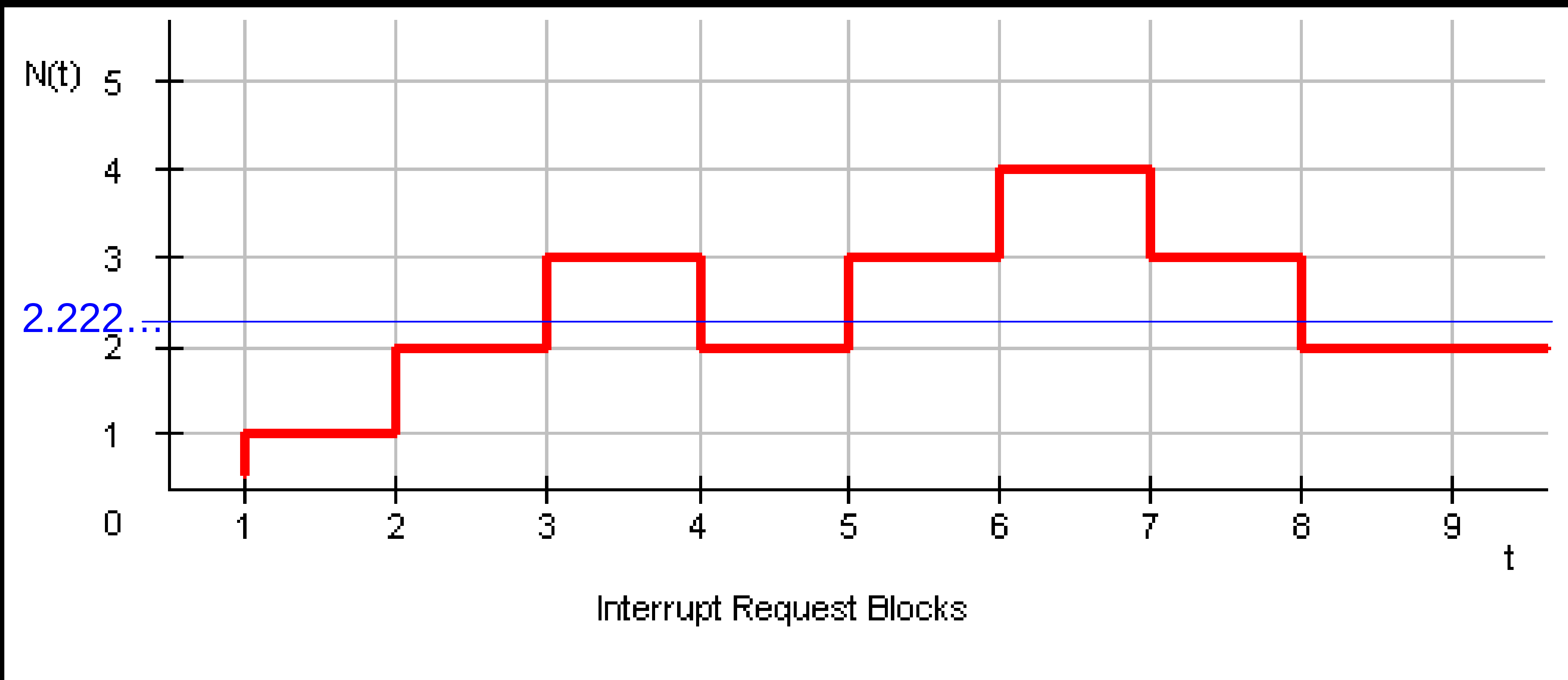
Example:



I/O Request Blocks and I/O Utilization

$$\overline{N}=\lim_{t?\ 8} N(t)$$

$$\overline{N}=\frac{(0+1+2+3+2+3+4+3+2)}{9}=2.222\ldots$$



I/O Request Blocks and I/O Utilization

$$\overline{N} = ?\, k \cdot f_k$$

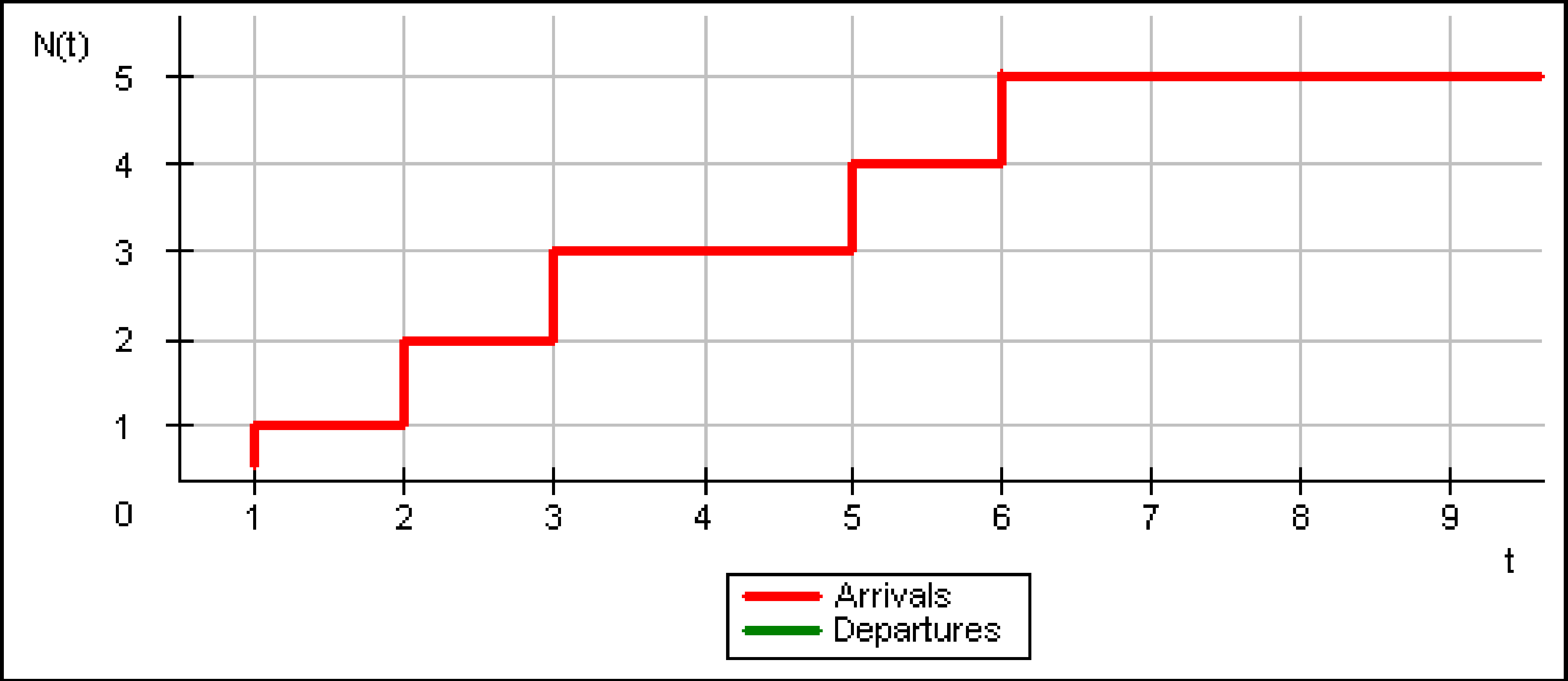
$$\overline{N} = \lim_{t ?\ 8} N(t)$$

N is the average number of IORB in the system

t is the time the system has been running

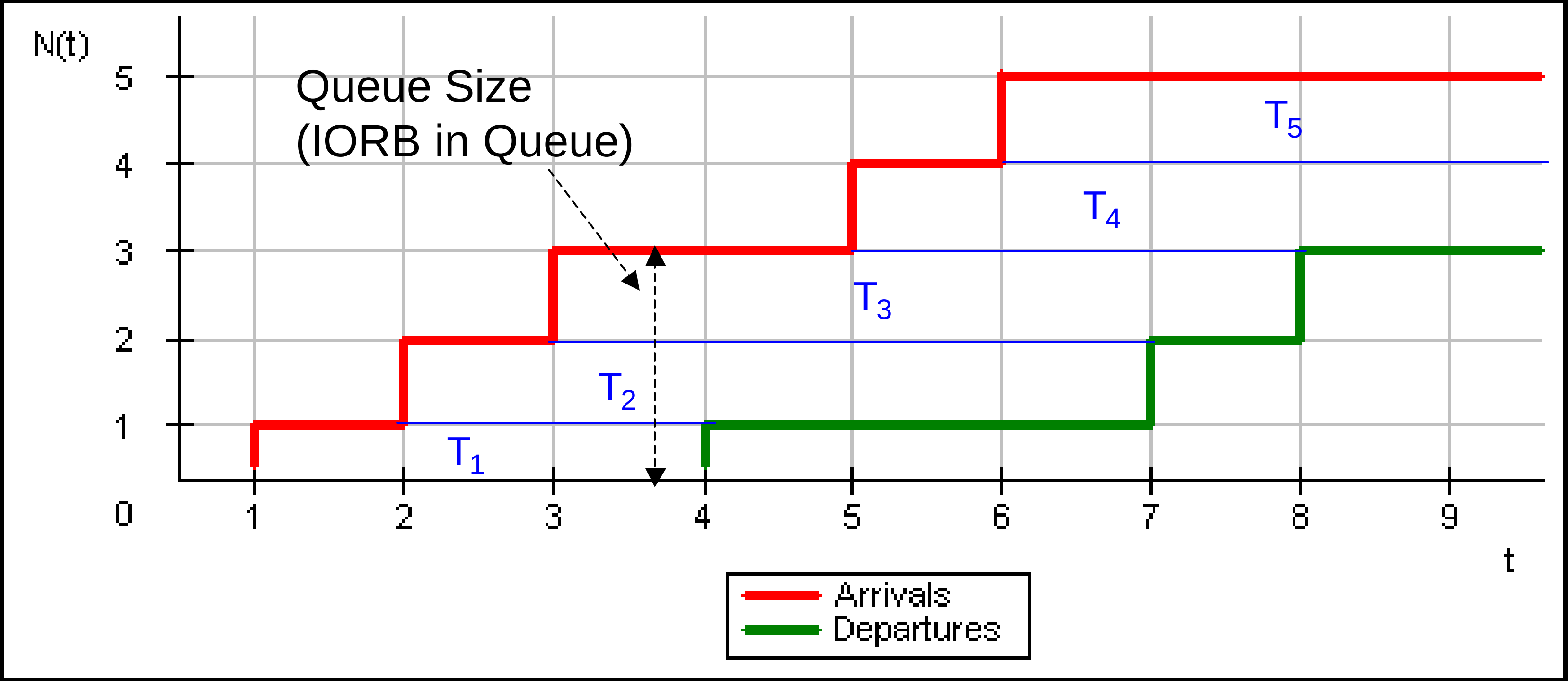
f_k is the fraction of time that there are k IORB in the system

I/O Request Blocks and I/O Utilization



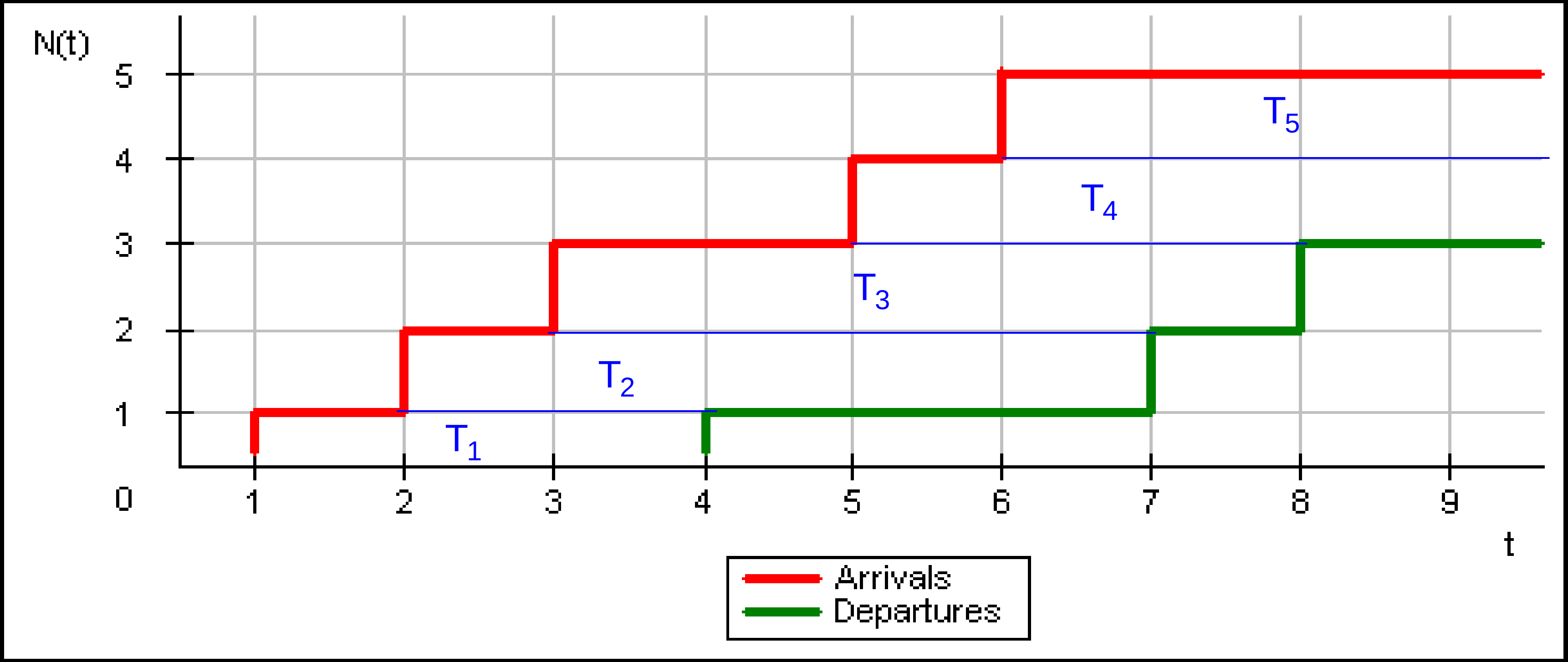
Showing only the Arrivals

I/O Request Blocks and I/O Utilization



T_i is the total service time for IORB i

I/O Request Blocks and I/O Utilization



The average service time:

$$\frac{\overset{5}{\underset{i=0}{?}} T_i}{\text{Max } N(t)} = \frac{T_1+T_2+T_3+T_4+T_5}{5}$$

I/O Request Blocks and I/O Utilization

Little’s Law

$$\overline{N}=?\cdot\overline{R}=\mu\cdot\overline{R}$$

$$\overline{N}=?\cdot T_{System}$$

$?=\mu$ in equilibrium

$?$ = the arrival rate = $\frac{A}{T}$

$\overline{R}$ = response Time

μ = service rate

I/O Request Blocks and I/O Utilization

Some Formulas:

$$? = \frac{A}{T}$$

$$U = \frac{?}{\mu}$$

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

$$\overline{R} = \frac{N}{?} = \frac{U}{1-U} \cdot \frac{1}{?} = \frac{?}{U(1-\frac{?}{U})} = \frac{1}{\mu-?}$$

$$\overline{R} = \frac{1}{\mu-?}$$

I/O Request Blocks and I/O Utilization

Average waiting time in the Queue

$$W_q=\overline{R}-\frac{1}{\mu}=\frac{1}{\mu-?}-\frac{1}{\mu}=\frac{U}{\mu-?}$$

The average waiting time in the queue is simply the average waiting time in the system minus the item currently being served.

I/O Request Blocks and I/O Utilization

Average number of IORB in the queue

By Little's Law:

$$\overline{N}_q = ? \cdot W_q = ? \cdot \frac{U}{\mu - ?} = \frac{? \cdot U}{\mu \cdot (1 - \frac{?}{U})} = \frac{U^2}{1-U}$$

Review:

$? = \frac{A}{T}$ $\overline{N} = \frac{U}{1-U}$ $W_q = \frac{U}{\mu - ?}$ $P_0 = 1-U$

$U = 1-P_0$

$U = \frac{?}{\mu}$ $\overline{R} = \frac{1}{\mu - ?}$ $\overline{N}_q = \frac{U^2}{1-U}$ $P_k = P_0 U^k$

$\overline{N}$=# of IORB in *system*

$\overline{N}_q$=# of IORB in *queue*

I/O Request Blocks and I/O Utilization

Example:

Requests arrive to the web-server at a rate of 30 requests per second. Each request takes 0.02 seconds on the average to be processed.

a. What is the fraction of time that (k=1, 2, 3, …) requests are found in the web-server?

b. R=?

I/O Request Blocks and I/O Utilization

Example:

a.

?=30

S=0.02s

$$\mu = \frac{1}{S} = 50 \text{ req/sec}$$

$$U = \frac{?}{\mu} = \frac{30}{50} = 3/5 = .6 < 1$$

U=60%

P_0=?

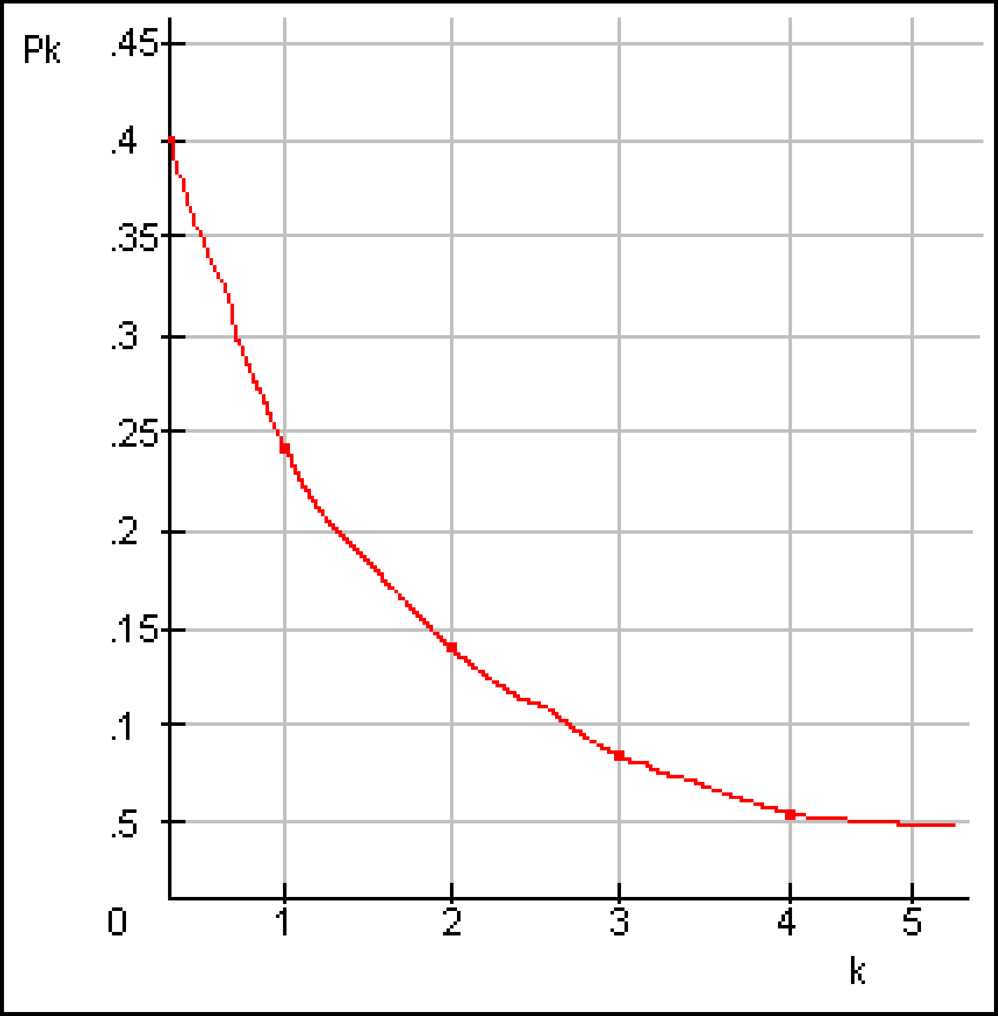
P_0 = (1-U) = 1-0.6 = 0.4 = 40%

I/O Request Blocks and I/O Utilization

Example:

a.

K	P_k	
1	0.24	$P_1=P_0U$
2	0.14	$P_2=P_0U^2$
3	0.086	$P_3=P_0U^3$
4	0.051	$P_4=P_0U^4$



I/O Request Blocks and I/O Utilization

Example:

b. one technique

$$\overline{N} = ? \cdot \overline{R} \qquad \overline{N} = \frac{60}{40} = \frac{3}{2}$$

$$\overline{R} = \frac{\overline{N}}{?} = \frac{3/2}{30} = .05 \text{ sec}$$

I/O Request Blocks and I/O Utilization

Example:

b. another technique (if you do not know $\bar{N}$)

$$\bar{N}=?\cdot\bar{R} \qquad \bar{N}=\frac{U}{1-U}$$

$$\bar{R}=\frac{\bar{N}}{?}=\frac{\frac{U}{1-U}}{?}=\frac{U}{?\cdot(1-U)}$$

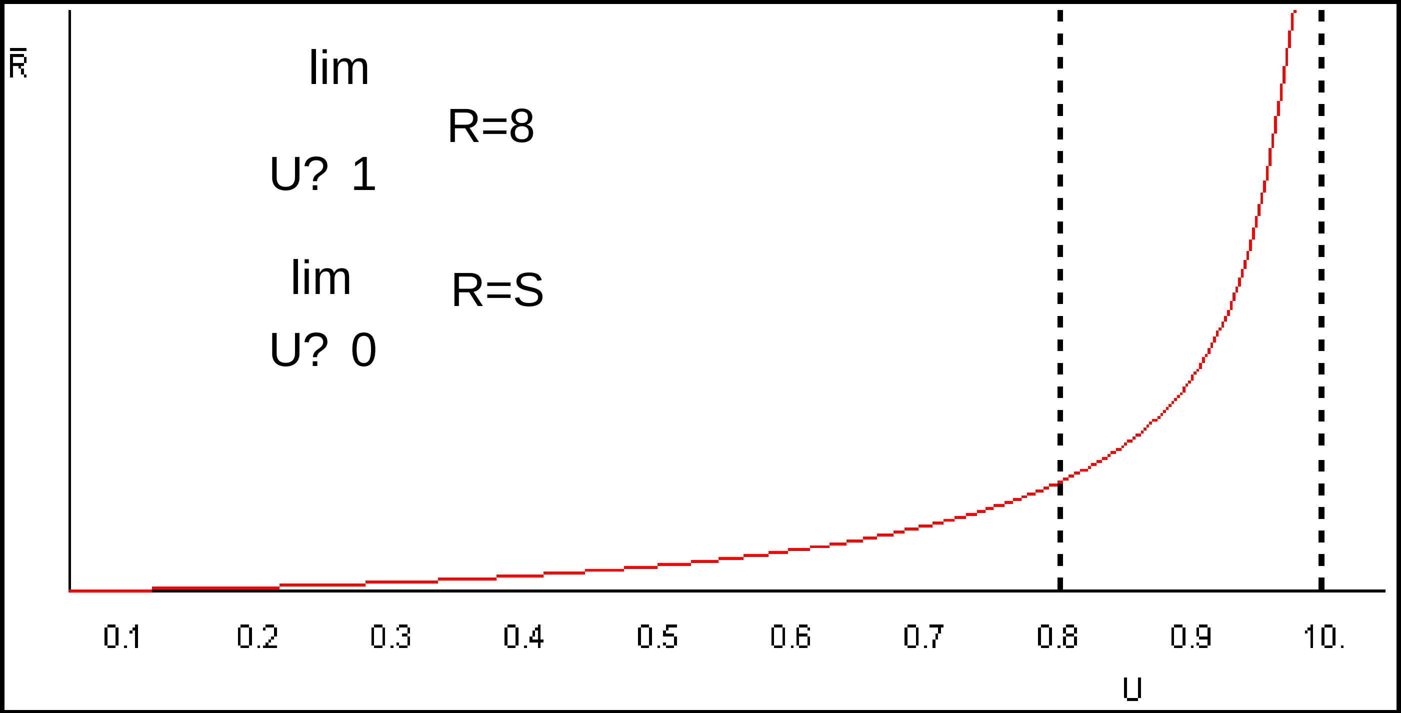
$$=\frac{\frac{?}{\mu}}{?\cdot(1-U)}=\frac{1}{\mu\cdot(1-U)}$$

$$=\frac{S}{1-U}=\frac{0.02}{1-0.6}$$

$$=0.05\text{ sec}$$

I/O Request Blocks and I/O Utilization

The ideal value for U



Ideal value for U is 0.8 (80%)