I/O Devices and Performance Measures

Lorn Miller
Steven Giovenco
I/O Request Blocks and I/O Utilization

I/ORB 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:

<table>
<thead>
<tr>
<th>T0</th>
<th>N(0)=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>N(1)=1</td>
</tr>
<tr>
<td>T2</td>
<td>N(2)=2</td>
</tr>
<tr>
<td>T3</td>
<td>N(3)=3</td>
</tr>
<tr>
<td>T4</td>
<td>N(4)=2</td>
</tr>
<tr>
<td>T5</td>
<td>N(5)=3</td>
</tr>
<tr>
<td>T6</td>
<td>N(6)=4</td>
</tr>
<tr>
<td>T7</td>
<td>N(7)=3</td>
</tr>
<tr>
<td>T8</td>
<td>N(8)=2</td>
</tr>
</tbody>
</table>
I/O Request Blocks and I/O Utilization

Example:

![Graph showing I/O Request Blocks and Utilization](image-url)
I/O Request Blocks and I/O Utilization

\[ \bar{N} = \lim_{t \to 8} N(t) \]

\[ \bar{N} = \frac{0+1+2+3+2+3+4+3+2}{9} = 2.222\ldots \]
I/O Request Blocks and I/O Utilization

\[ \bar{N} = \sum_{k} k \cdot f_k \]

\[ \bar{N} = \lim_{t \to \infty} N(t) \]

\( \bar{N} \) 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
$T_i$ is the total service time for IORB $i$
The average service time:

\[
\frac{\sum_{i=0}^{5} T_i}{\text{Max } N(t)} = \frac{T_1 + T_2 + T_3 + T_4 + T_5}{5}
\]
I/O Request Blocks and I/O Utilization

**Little’s Law**

\[
\bar{N} = \bar{R} \cdot \mu = \mu \cdot \bar{R} \\
\bar{N} = \bar{R} \cdot T_{\text{System}}
\]

\(\bar{N}\) = \(\bar{R}\) in equilibrium
\(\bar{R}\) = the arrival rate = \(\frac{A}{T}\)
\(\bar{R}\) = response Time
\(\mu\) = service rate
I/O Request Blocks and I/O Utilization

Some Formulas:

\[
? = \frac{A}{T} \\
U = \frac{?}{\mu} \\
\bar{N} = \frac{U}{1-U} \\
\bar{R} = \frac{N}{?} = \frac{U}{1-U} \cdot \frac{1}{?} = \frac{?}{U(1-\frac{?}{U})} = \frac{1}{\mu-?} \\
\bar{R} = \frac{1}{\mu-?}
\]
I/O Request Blocks and I/O Utilization

Average waiting time in the Queue

\[ W_q = R - \frac{1}{\mu} = \frac{1}{\mu - ?} - \frac{1}{\mu} = U \]

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:

\[ \bar{N}_q = \rho \cdot W_q = \frac{\rho \cdot U}{\mu - \rho} = \frac{\rho \cdot U}{\mu \cdot (1 - \frac{\rho}{\mu})} = \frac{U^2}{1 - U} \]

Review:

\[ \rho = \frac{A}{T} \]
\[ N = \frac{U}{1 - U} \]
\[ W_q = \frac{U}{\mu - \rho} \]
\[ P_0 = 1 - U \]
\[ U = 1 - P_0 \]
\[ R = \frac{1}{\mu - \rho} \]
\[ \bar{N}_q = \frac{U^2}{1 - U} \]
\[ P_k = P_0 U^k \]

\( \bar{N} = \# \text{ of IORB in system} \)

\( \bar{N}_q = \# \text{ of IORB in queue} \)
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.

\[ \tau = 30 \]
\[ S = 0.02 \text{s} \]

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

\[ U = \frac{\tau}{\mu} = \frac{30}{50} = 0.6 < 1 \]

\[ U = 60\% \]

\[ P_0 = ? \]

\[ P_0 = (1 - U) = 1 - 0.6 = 0.4 = 40\% \]
I/O Request Blocks and I/O Utilization

Example:

a.

<table>
<thead>
<tr>
<th>K</th>
<th>P_k</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.24</td>
</tr>
<tr>
<td>2</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>0.086</td>
</tr>
<tr>
<td>4</td>
<td>0.051</td>
</tr>
</tbody>
</table>
I/O Request Blocks and I/O Utilization

Example:

b. one technique

\[ \bar{N} = \overline{R} \]

\[ \bar{N} = \frac{60}{40} = \frac{3}{2} \]

\[ \overline{R} = \frac{\bar{N}}{30} = \frac{3/2}{30} = 0.05 \text{ sec} \]
I/O Request Blocks and I/O Utilization

Example:

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

\[
\bar{N} = \bar{R} \cdot \bar{N} = \frac{U}{1-U}
\]

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

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

\[
\bar{S} = \frac{0.02}{1-U} = \frac{1}{1-0.6}
\]

\[
= 0.05 \text{ sec}
\]
I/O Request Blocks and I/O Utilization

The ideal value for $U$ is 0.8 (80%)