Solution of the center with priority discipline

The priority disciplines are those where users are served based on fixed priorities assigned to them.

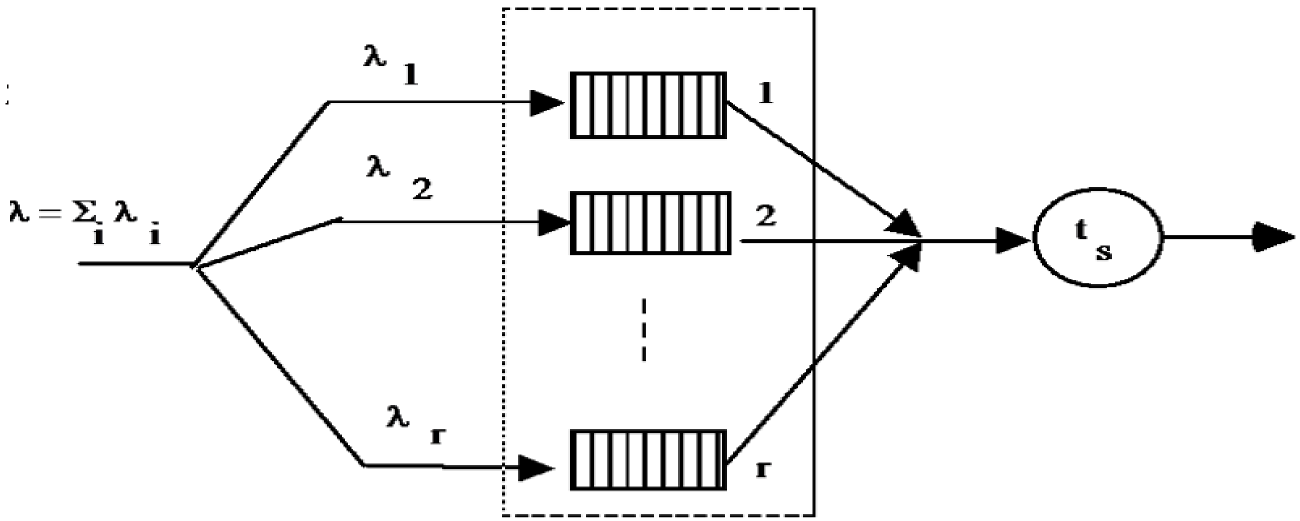
These disciplines can be *with preemption* (denoted WP) or with *No-preemption* (denoted NP)

Solution of the center with priority discipline

In the *no preemption* case, the arrival of a priority job doesn't interrupt the service in progress

Instead, in the case *with preemption*, the arriving user who has greater priority compared to the one currently in service, can interrupt this execution to start his own

Solution of the center with priority discipline



Single center and priority queue model

Solution of the center with priority discipline

The waiting queue is only conceptually divided into *r* queues of several priorities, where the maximum priority jobs wait in the level 1 queue, while the minimum priority jobs wait in *r* queue

As already said, priority is said to be ***abstract*** if users are divided among the various levels based on the criteria that doesn't depend on the service time they request to the center

Solution of the center with priority discipline

Priority is instead said to be service-***time dependent***, if users who ask for a a t_s included in a certain defined interval enter level 1, while those who ask for a $t_{s,}$ included in a different interval enter at level 2 , and so on.

Abstract Priority No Preemption (APNP) Discipline

The parameters relating to class k, $1 \leq k \leq r$, are

- λ_k — mean frequency of arrivals at level k
- t_{sk} — service time of a class k user
- $E(t_{sk})=1/\mu_k$ — mean of the t_{sk}
- $\sigma^2(t_{sk})$ — variance of the t_{sk}
- $\rho_k = \lambda_k / \mu_k$ — center utilization by users of level k

Abstract Priority No Preemption (APNP) Discipline

- The discipline is FIFO inside each level
- Users in class k are served only if there aren't any in the classes (1, 2, ..., k-1)
- If a service to a k priority user is in progress, and a user arrives in one of the higher classes (1, 2, ..., k-1), the center ends the service in progress (no preemption) and then moves on to serve the arrival with greater priority

Abstract Priority No Preemption (APNP) Discipline

The mean waiting time of a user u_k belonging to class k, denoted as $E(t_{wk})_{APNP}$, depends on three factors, called **waiting components**:

I. Mean time needed to complete the service in progress
II. Mean time needed to serve the users already present in the queues with greater priority and in the same arrival queue
III. Mean time needed to serve the users in classes 1, 2, .., k-1 that arrived after the arrival of u_k but before it receives the service.

According to what said on the relationship between mean waiting time and mean seen remaining time, such quantities can be written as follows:

Abstract Priority No Preemption (APNP) Discipline

I. A quantity equal to

$$\frac{\lambda}{2} E(t_s^2) \qquad \text{with} \quad \lambda = \sum_{k=1}^{r} \lambda_k$$

II. A quantity proportional to according to the law $\dfrac{1}{1 - \sum_{i=1}^{k} \rho_i}$ $\qquad \sum_{i=1}^{k} \rho_i$

Abstract Priority No Preemption (APNP) Discipline

III. A quantity proportional to $\sum_{i=1}^{k-1} \rho_i$

according to the law $\dfrac{1}{1 - \sum_{i=1}^{k-1} \rho_i}$

These quantities combine in a multiplicative way and give rise to

$$E(t_{w_k})_{APNP} = \frac{\frac{\lambda}{2} E(t_s^2)}{(1 - \sum_{i=1}^{k} \rho_i)\ (1 - \sum_{i=1}^{k-1} \rho_i)} \qquad (1)$$

Abstract Priority No Preemption (APNP) Discipline

We can observe that if the first waiting component is null, then the classes are all empty on the arrival of user u_k and therefore all of $E(t_{wk})_{APNP}$ is null.

Abstract Priority No Preemption (APNP) Discipline

By applying (1) to two contiguous classes k and k+1, it is easy to verify the effect of the discipline, in that the following relation will be true:

$$E(t_{wk})_{APNP} \leq E(t_{wk+1})_{APNP}$$

In other words, the users in class with greater priority experiment a lower mean waiting time

Abstract Priority No Preemption (APNP) Discipline

From relation (1), if we apply Little's law, i.e. multiplying by λ_k, we obtain the mean length of the queue k $(E(w_k)_{APNP})$

Where if we sum up ρ_k it is possible to obtain the mean population of level k $(E(q_k)_{APNP})$

With Little, dividing by $\lambda_{k,}$ we obtain the response time at level k $(E(t_{qk})_{APNP})$

Abstract Priority No Preemption (APNP) Discipline

In regards the center as a whole, the ***mean overall waiting time*** is obtained by

$$E(t_w)_{APNP} = E\,[E(t_{wk})_{APNP}\,] = \sum_{k=1}^{r} p_k\, E(t_{wk})_{APWP}$$

With $p_k = \frac{\lambda_k}{\lambda}$ the probability of entering in level k

Time priority no preemption discipline (TPNP)

TPNP is a discipline based on the service time t_s and without preemption, or in other words, the arrivals which ask for a shorter service time are entered in classes with greater priorities

With two priority classes called *h* and *k*, if $h < k$ then it will also be $E(t_{sh}) < E(t_{sk})$

Time priority no preemption discipline (TPNP)

Assume $t_s \in [0, D]$ and $0 = d_0 < d_1 < d_2 < \ldots < d_r = D$

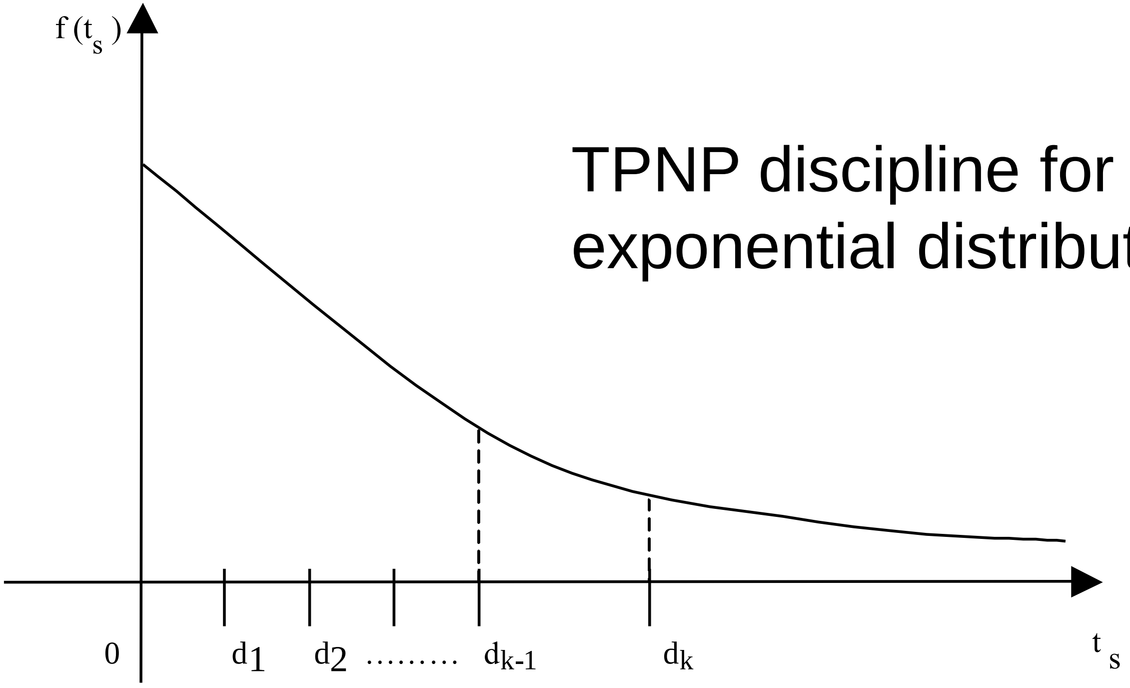
We therefore define the user of class k if the service time he asks for is $t_s \in [d_{k-1}, d_k]$

- In each class, the service discipline is FIFO
- The time $E(t_{w_k})_{TPNP}$ can be calculated based on the formula for the calculation of $E(t_{w_k})_{APNP}$, by suitably interpreting the λ_k and the $E(t_{sk})$, k=1, …, r

Time priority no preemption discipline (TPNP)

In fact, while the arrival rates λ_k for APNP are arbitrarily set and $E(t_{sk}) = E(t_s)$, in TPNP the values of λ_k and $E(t_{sk})$ depend on the form of the distribution $f(t_s)$ of the service time t_s and on how the intervals $(d_{k-1}, d_k]$ are chosen

Time priority no preemption discipline (TPNP)



TPNP discipline for the exponential distribution case

Time priority no preemption discipline (TPNP)

Given this division into intervals in the exponential density $f(t_s)$ case, the value $E(t_{sk})$ is the mean of times which belong to the interval $(d_{k-1}, d_k]$, and is therefore obtainable as

$$E(t_{s_k}) = \int_{d_{k-1}}^{d_k} t_s \, f_k^n(t_s) dt_s$$

Time priority no preemption discipline (TPNP)

where $f_k^n(t_s) = \dfrac{f_k(t_s)}{F(d_k) - F(d_{k-1})}$

represents the portion of $f(t_s)$ relating to interval k, normalized in a way that the underlying area between d_{k-1} and d_k is unitary

Time priority no preemption discipline (TPNP)

In the same way, λ_k is proportional to the relation between the portion of area under the curve d_{k-1} and d_k and the total area. Because the total area is equal to 1, there results that λ_k is that part of λ proportional to $F(d_k)$ - $F(d_{k-1})$

In other words, $\lambda_k = \lambda\left(F(d_k) - F(d_{k-1})\right)$

Time priority no preemption discipline (TPNP)

By starting from the expression $E(t_{wk})_{APNP}$ it is possible to obtain an expression for $E(t_{wk})_{TPNP}$ in the form

$$E(t_{w_k})_{TPNP} = \frac{\frac{\lambda}{2} E(t_s^2)}{\left(1 - \lambda \int_0^{d_k} t_s \, dF(t_s)\right)\left(1 - \lambda \int_0^{d_{k-1}} t_s \, dF(t_s)\right)}$$

Time priority no preemption discipline (TPNP)

It is also possible to convince ourselves that, with all factors equal, the following relation is true

$$E(t_{wk})_{TPNP} \leq E(t_{wk})_{APNP}$$

Below are the following general conditions

- Appropriate form (e.g. exponential) of $f(t_s)$
- Intervals $(d_{k-1}, d_k]$ in order to ensure a suitable order of the series of λ_i and $E(t_{si})$

Time priority no preemption discipline (TPNP)

It is now possible to express the ***mean overall waiting time*** $E(t_w)_{TPNP}$ by starting from the definition of the mean of means and thus

$$E(t_w)_{TPNP} = E\,[E(t_{w_k})_{TPNP}] = \sum_{k=1}^{r} p_k E(t_{w_k})_{TPNP}$$

With $p_k = \dfrac{\lambda_k}{\lambda}$ the probability of entering in level k and

for what was said $p_k = F(d_k) - F(d_{k-1})$

Time priority no preemption discipline (TPNP)

It is therefore possible to write $E(t_w)_{TPNP}$ as

$$E(t_w)_{TPNP} = \frac{\lambda}{2} E(t_s^2) \sum_{k=1}^{r} \frac{F(d_k) - F(d_{k-1})}{\left(1 - \lambda \int_0^{d_k} t_s dF(t_s)\right)\left(1 - \lambda \int_0^{d_{k-1}} t_s dF(t_s)\right)}$$

Time priority no preemption discipline (TPNP)

It is possible to state that

$$E(t_w)_{TPNP} \le E(t_w)_{APNP}$$

Demonstrable by applying the definition of mean of means seen above

This result allows us to state that even the overall mean waiting time in the TPNP case improves compared to APNP case.

Time priority no preemption discipline (TPNP)

In a similar way as to how we calculate the overall mean waiting time, it is possible to get the overall mean $E(t_q)_{TPNP}$ of response time t_q.

In fact the following is true

$$E(t_q)_{TPNP} = E(E(t_{q_k}))TPNP = \sum_{k=1}^{r} p_k E(t_{q_k})TPNP$$

where:

$$E(t_{q_k})_{TPNP} = E(t_{w_k})_{TPNP} + E(t_{s_k})_{TPNP}$$

Time priority no preemption discipline (TPNP)

On the other hand, because the mean of the means of service times t_s is the same in TPNP and APNP, the comparison between the overall residence times TPNP and APNP is reduced to the comparison between the mean waiting times $E(t_w)$, where we can state that

$$E(t_q)_{TPNP} \leq E(t_q)_{APNP}$$

Shortest Processing Time First (SPTF) Discipline

The TPNP can still be improved if we eliminate the abstract residual component which remains in FIFO inside each priority level

This is obtainable by introducing a free number of levels as the number of waiting jobs. One level for each job.

Shortest Processing Time First (SPTF)

In this way, each class contains one job, and priority is always given to the job which requires the lowest service time

The ***overall average mean waiting time*** is obtainable by the expression of $E(t_w)_{TPNP}$ making $r \to \propto$.

From which

$$E(t_w)_{SPTF} = \frac{\lambda}{2} E(t_s^2) \int_0^{\infty} \frac{dF(t_s)}{\left(1 - \lambda \int_0^{t_s} t \, dF(t)\right)^2}$$

Shortest Processing Time First (SPTF)

By elaborating on what was said for the TPNP case, we can state that

$$E(t_w)_{SPTF} \le E(t_w)_{TPNP} \le E(t_w)_{APNP}$$

In other words, between the no preemption priority disciplines, the SPTF is the best in terms of mean waiting time

Abstract Priority with preemption discipline (APWP)

This discipline allows an arriving class k user, to interrupt the user currently in service, if the latter is of lower priority, therefore of class k+1, .., r

Abstract Priority with preemption discipline (APWP)

The interrupted user doesn't go back to the tail of its queue but remains at the head ready to restart the execution according to the two ways that have already been discussed

1. With loss (denoted *lo*)
2. Lossless (denoted *ls*)

Abstract Priority with preemption discipline (APWP)

In the *lossless* case, the mean waiting time of a user who enters class k, $E(t_{wk})_{APWP}$, is composed, like APNP, of three components

1. Mean time needed to complete the current non interruptible service
2. The mean time needed to serve the users already present in the queues with greater priority and those in the same queue during the user arrival, therefore the users present in categories 1, 2,..., k
3. The mean time needed to serve the users arrived in classes 1, 2, .., k-1 while ours was waiting

Abstract Priority with preemption discipline (APWP)

While the second and third component take on similar values to the APNP case, the first is the function of only the first k priority levels and it is expressed with the

$$\left(\frac{1}{2}\sum_{i=1}^{k}\lambda_i\right)E(t_s^2)$$

Abstract Priority with preemption discipline (APWP)

Like in the APNP case, the three components combine in a multiplicative relation amongst themselves, and thus

the mean waiting time of a class k user is

$$E(t_{w_k})_{APWP\,ls} = \frac{\left(\frac{1}{2}\sum_{i=1}^{k}\lambda_i\right) E(t_s^2)}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)}$$

Abstract Priority with preemption discipline (APWP)

In the case where the discipline is lossless, the service time to give to the interrupted user is again equal to the whole service time

This causes a system overload. We can demonstrate that $E(t_{wk})_{APWPlo} \geq E(t_{wk})_{APWPls}$ in that the user who arrives finds a more congested system

Abstract Priority with preemption discipline (APWP)

By limiting ourselves to the *lossless (ls) case,* we can observe that the mean service time experimented by each user of class k ($E(tw_k)_{APWPls}$), gets longer in relation to the fact that the user in service can be interrupted by jobs which arrive and have greater priority

Its service time therefore virtually gets longer in proportion to the number of times that it is interrupted. In essence, a user undertakes a longer "virtual" t_s the shorter its priority is.

Virtual Service Time

The virtual service time, observed by the class k user, can be expressed as

$$E(t_{s_k virt})_{APWP_{ls}} = \frac{E(t_s)}{1 - \sum_{i=1}^{k-1} \rho_i}$$

It therefore results in $E(t_{s_k virt})_{APWP_{ls}} \geq E(t_s)$

or, each class k user will experiment a greater mean service time compared to the mean service time he asks for, due to the preemption effect.

Abstract Priority with preemption discipline (APWP)

By comparing the relations, it is possible to convince ourselves that

$$E(t_{wk})_{APWPls} \leq E(t_{wk})_{APNP}$$

And also $E(t_{s_{kvirt}})_{APWP_{ls}} \geq E(t_{s_k})_{APNP}$

being $E(t_{sk})_{APNP} = E(t_s)$

In other terms, in $APWP_{ls}$ the mean waiting time per class is shorter than in the APNP (thanks to preemption), while the mean service time is worse (due to service interruption).

Abstract Priority with preemption discipline (APWP)

Because, in $APWP_{ls}$ of class k, the service time and waiting time have contrasting trends, nothing general can be asserted for the mean response time of class k, sum of the mean waiting time and the mean virtual service time, as opposed to the APNP

Abstract Priority with preemption discipline (APWP)

By calculating the mean of the means of *r* priority levels, we obtain a relation for the mean overall waiting time

$$E(t_W)_{APWP_{ls}} = E\,[\,E(t_{w_k})_{APWP_{ls}}\,] = \sum_{k=1}^{r} p_k\, E(t_{w_k})_{APWP_{ls}}$$

Abstract Priority with preemption discipline (APWP)

Similarly we can obtain an expression for the mean overall response time

$$E(t_q)_{APWPls} = E\,[E(t_{qk})_{APWPls}] = \sum_{k=1}^{r} p_k E(t_{q_k})_{APWPls}$$

Abstract Priority with preemption discipline (APWP)

We can gather (by making calculations for the two-level priority case) that the lossless with preemption discipline can't have any effects on the overall response time with respect to the APNP and KP and therefore, we can write

$$E(t_q)_{APWPls} = E(t_q)_{APNP} = E(t_q)_{KP}$$

Time Priority with preemption (TPWP) discipline and Shortest Remaining Processing Time First (SRPTF)

Let's now consider a time dependent discipline with preemption (TPWP), limited to the *lossless* *(ls)* case.

The users who request a shorter service time enter inside the levels of greater priority and can interrupt the services of lower priority

Time Priority with preemption (TPWP) discipline and Shortest Remaining Processing Time First (SRPTF)

Three cases can be distinguished

1. TPWPsimple: discipline in which the interrupted user remains in the class where he entered with his remaining service time (thus lossless)
2. TPWPsemipreemption: discipline in which the interrupted user (of class k+1, … ,r) goes in the class corresponding to its remaining service time (therefore still a lossless discipline) and thus, the priority rises
3. SRPTF: Shortest Remaining Processing Time First discipline, which is a semi-preemption with a infinite number (*r*) of classes

Time priority with preemption discipline and SRPTF

As already was done for the Time Priority no preemption, assume ts $\in$ [0, D] and $0 = d_0 < d_1 < d_2 ... < d_r = D$

We define the *class k* user if the service time he asks for is $t_s \in (d_{k-1}, d_k]$

It is assumed the discipline internal at each class remains FIFO

Time priority with preemption discipline and SRPTF

Having made these assumptions, we can state that for the $TPWP_{simple}$ the mean waiting time of level k, $E(t_{wk})_{TPWPsimple}$, can be calculated based on the APWPs formula, by suitably interpreting the parameters λ_i and $E(t_{si})$, for i = 1,..., k (which appear in the numerator and in the ρ_i of the denominator)

You can refer to the textbook for further information

Time priority discipline with preemption and SRPTF

In regards the $TPWP_{semipreemption}$ discipline, the mean for level $E(t_{wk})_{TPWPsemipreemption}$ can be expressed by starting from the formula given for $APWP_{ls}$, by transforming it like in the TPNP case

The difference in this discipline is in the calculation of the numerator, or in other words, in the remaining time expression

(follow the book for details)

Time priority discipline with preemption and SRPTF

The mean waiting time for a class k user is therefore expressed by

$$E(t_{w_k})_{TPWPsemipreemption} = \frac{\lambda}{2} \frac{\int_0^{d_k} t_s^2 dF(t_s) + (1 - F(d_k))d_k^2}{(1 - \lambda \int_0^{d_k} t_s dF(t_s))(1 - \lambda \int_0^{d_{k-1}} t_s dF(t_s))}$$

Time priority discipline with preemption and SRPTF

Moving ahead to the overall mean waiting time, we can write the following equation

$$E(t_W)_{\text{TPWP semipreemption}} = E\,[E(t_{W_k})_{\text{TPWP semipreemption}}] = \sum_{k=1}^{r} p_k E(t_{w_k})_{TPWP semipreemption}$$

With $p_k = \dfrac{\lambda_k}{\lambda}$ the probability of entering in level k given

as usual by $p_k = F(d_k) - F(d_{k-1})$

Time priority discipline with preemption and SRPTF

We can therefore write the overall mean waiting time as

$$E(t_w)_{TPWPsemipreemption} = \frac{\lambda}{2}\sum_{k=1}^{r}\frac{(F(d_k)-F(d_{k-1}))(\int_0^{d_k} t_s^2\, dF(t_s) + (1-F(d_k))d_k^2)}{(1-\lambda\int_0^{d_k} t_s\, dF(t_s))(1-\lambda\int_0^{d_{k-1}} t_s\, dF(t_s))}$$

Look at the book for the relations between $E(t_{wk})$ and various $E(t_w)$

Time priority discipline with preemption and SRPTF

The *Shortest Remaining Processing Time First* discipline can be seen as an extension of the $TPWP_{semipreremption}$ discipline by making the number *r of* priority levels tend to infinity.

Thus by operating the limit of the $E(T_w)_{semipreremption}$ formula one may derive the $E(t_w)_{SRPTF}$ one

Time priority discipline with preemption and SRPTF

In the *SRPTF* case we therefore obtain

$$E(t_w)_{SRPTF} = \frac{\lambda}{2} \int_0^{\infty} \frac{\int_0^{t_s} t^2 \, dF(t) + (1 - F(t_s))\, t_s^2}{(1 - \lambda \int_0^{t_s} t \, dF(t))^2} \, dF(t_s)$$

TPWP and SRPTF disciplines

We observe that with the SRPTF discipline, we eliminate a last abstraction element still present in the SPTF, i.e. the fact that in this latter the priority is denied to a user in arrival with a shorter service request compared to the remaining service in progress, which thus, forces a FIFO between the two.

TPWP and SRPTF disciplines

We must expect that the SRPTF is better than SPTF, in terms of mean waiting time

By comparing the suitable relations (see the book for the formulae) it is easy to convince ourselves that the following relations are valid:

$$E(t_w)_{SRPTF} \leq E(t_w)_{SPTF} \overset{?}{\leq} E(t_w)_{TPWP\,semipreemption}$$

$$\leq E(t_w)_{TPWP} \leq E(t_w)_{APWP} = E(t_w)_{KP}$$

TPWP and SRPTF disciplines

Where the $\stackrel{?}{\leq}$ relation indicates that it isn't always true that

$$E(t_w)_{SPTF} \leq E(t_w)_{TPWPsemipreemption}$$

(consult the book for details on the comparison)

TPWP and SRPTF disciplines

Nevertheless, from this comparison, we deduct that the time priority disciplines, even if very easy, provide better results in terms of overall waiting time compared to the abstract priorities, APNP, which behave like the ones without priority (FIFO) in respect to the overall waiting time

TPWP and SRPTF disciplines

To complete the TPWPsemipreemption and SRPTF disciplines, we can respectively express the overall mean response time of both disciplines,

$$E(t_q)_{PTCPsemipreemption} = E(t_w)_{PTCPsemipreemption} +$$

$$+\sum_{k=1}^{r}\left[\frac{\int_{d_{k-1}}^{d_k} t_s f(t_s)dt_s}{1-\lambda\int_0^{d_{k-1}} t_s f(t_s)dt_s} + (F(d_k)-F(d_{k-1}))\left(-\frac{d_{k-1}}{1-\lambda\int_0^{d_{k-1}} t_s f(t_s)dt_s} + \sum_{i=1}^{k-1}\frac{d_i - d_{i-1}}{1-\lambda\int_0^{d_{i-1}} t_s f(t_s)dt_s}\right)\right]$$

$$E(t_q)_{SRPTF} = E(t_w)_{SRPTF} + \int_0^{\infty}\left[\int_0^{t_s}\frac{dt_s}{1-\lambda\int_0^{d_{i-1}} t_s f(t_s)dt_s}\right] dF(t_s)$$

TPWP and SRPTF disciplines

To demonstrate what was said until now, we can observe the following table as a comparative example of the various disciplines based on mean response time $E(t_q)$

TPWP and SRPTF disciplines

Discipline	*E(t_q)*	*Formula used*
FIFO	5.0	(6.12) e (6.3c)
2/APNP	5.0	(6.13’)
2/TPNP	3.3	(6.16)
SPTF	2.9	(6.17)
2/TPWPsemipreemption	2.8	(6.25)
SRPTF	2.3	(6.26)

The resolved case is of an exp center with $E(t_s) = 1$ and $\lambda = 0.8$ ($\rho = 0.8$)

(consult the book for the comments on the results obtained by the system solution)