On systems with shared resources and optimal switching strategies

Maxim Finkelstein (FinkelM.SCI@ufs.ac.za)
ON SYSTEMS WITH SHARED RESOURCES AND OPTIMAL SWITCHING STRATEGIES

Maxim Finkelstein

Department of Mathematical Statistics
University of the Free State
PO Box 339, 9300 Bloemfontein, Republic of South Africa
E-mail: FinkelM.SCI@ufs.ac.za

and

Max Planck Institute for Demographic Research,
Konrad-Zuse-Strasse 118057 Rostock, Germany,

ABSTRACT

Simple series systems of identical components with spare parts are considered. It is shown that the cumulative distribution function of a system failure time tends to a step function as the number of components increases and resources can be shared. An example of ‘continuous resources’ is also described. The time-sharing strategy for standby systems is investigated. It is proved that an optimal rule for a system of standby components with increasing failure rates is the single switching performed at $a = t/2$, where $t$ is a mission time.

Keywords: resource, spare part, convergence in distribution, aging, optimal switching strategy.

1. INTRODUCTION

In this paper, we consider performance of systems with initial (at $t = 0$) resource of some kind. The simplest example of a resource is the spare parts for an operating component (standby reserve). Each failure of an operating component is ‘repaired’ by the substitution of the spare one. The failure of a system occurs when there are no available spare parts for substitution. In a general setting, the run out resource increases in time and the failure occurs when it reaches some deterministic or random value. If a system consists of a number of components with resources that can be shared, then sharing increases reliability of such systems. When a large number of identical components share resources, some interesting facts can be observed. In Sections 2 and 3, specifically, we show that, as the number of components increases, the failure time of a series system tends to be ‘less random’ in the sense to be defined later. This means that the corresponding reliability of a system under certain assumptions can be extremely high (close to 1) in the interval $[0,m]$, where $m$ is the mean lifetime of a component, and then falls down sharply to 0 outside this interval. We believe that this property can be useful in designing reliable systems.

When considering performance of standby systems with aging components, the following question arises: will the switching from one component to another prior to its failure increase the performance characteristics of the whole system? We understand here aging as increasing of the corresponding failure rate or as decreasing of a performance characteristic of a component and show in Sections 4 and 5 that such simple switching strategies exist. Specifically, it turns out that switching at $a = t/2$,
where \( t \) is a mission time, maximizes the probability of the system’s performance without failures of components in \([0,t]\). The strategy that maximizes the mean time to this event is also suggested.

2. SHARED SPARE PARTS

Consider a series system of \( n \) i.i.d. components with the Cdf \( F_c(t) \). Assume that there are \( n_i \geq 0 \) spare parts (‘cold’ standby) for the component \( i, 1 \leq i \leq n \). Assume that the replacement of the failed component is instantaneous. The system fails when one of the components fails and there are no more spare parts allocated to this component. The survival function (\( \bar{F} \equiv 1 - F \)) of our series system is defined in this case as

\[
\bar{F}(t) = \prod_{i=1}^{n} \bar{F}_c^{(n)}(t),
\]

where \( \bar{F}_c^{(n)}(t) \) is the \( n \)-fold convolution of \( F_c(t) \) with itself and \( \bar{F}_c^{(0)}(t) \equiv F_c(t) \).

It is clear from obvious probabilistic considerations [10] that the survival probability increases if spare parts can be shared, i.e.,

\[
\bar{F}_s(t) \geq \bar{F}(t), t \geq 0,
\]

where \( \bar{F}_s(t) \) denotes survival probability of the system with shared resources. For the specific exponential case, this obvious fact can be illustrated by the following remarkable asymptotic considerations:

**Example 1.** Let \( F_c(t) = 1 - \exp\{\lambda t\} \) and \( n_i = m, i = 1,2,\ldots,n \). Then

\[
\bar{F}_c^{(m)}(t) = \exp\{-\lambda t\} \sum_{0}^{m} \frac{(\lambda t)^i}{i!}
\]

and equation (1) becomes

\[
\bar{F}(t) = (\bar{F}_c^{m}(t))^n.
\]

It is clear that for every fixed \( t > 0 \) and \( m \geq 0 \) the survival function vanishes: \( \bar{F}(t) \to 0 \) as \( n \to \infty \). On the other hand, assume now that resources are shared. Then the corresponding survival probability is

\[
\bar{F}_s(t) = \exp\{-n\lambda t\} \sum_{0}^{nm} \frac{(n\lambda t)^i}{i!}
\]

Obtaining the limit values for \( m \geq 1 \) and every fixed \( t > 0 \) in this case is not so straightforward. Let for simplicity \( m = 1 \). As equation (4) defines the Erlangian Cdf (the sum of \( n + 1 \) i.i.d. exponentially distributed random variables with mean \( 1/n\lambda \)), the central limit theorem can be applied, which gives for \( n \to \infty \):
\[ F_s(t) \to \Phi \left( \frac{t - \frac{n + 1}{n\lambda}}{\sqrt{n + 1}/n\lambda} \right) \approx \Phi \left( \frac{t - 1/\lambda}{1/\sqrt{n}\lambda} \right) = N \left( \frac{1}{\lambda}, \frac{1}{n\lambda^2} \right) \]  

(5)

where, as usually, \( \Phi(\cdot) \) denotes the Cdf of the standard normal distribution \( N(0,1) \). As the variance tends to 0 with \( n \to \infty \), relationship (5) means that \( F_s(t) \) converges uniformly to the step function:

\[ F_s(t) \to I(t) = \begin{cases} 0, & t < 1/\lambda \\ 1, & t \geq 1/\lambda \end{cases} \]

(6)

which, in fact, means that the time to failure of our system with shared spares converges in distribution to a non-random \( 1/\lambda \). This conclusion surprisingly differs from the case of the not shared resources. It is obvious that this result is valid for \( m > 1 \) and the rate of convergence in (5)-(6) increases with \( m \). It is also clear that the assumption of identical components is essential for sharing of resources, although a generalization to the case when the spare parts for different components are non-identical but interchangeable can be also considered (see the next section).

**Remark 1.** Relationships (6) is a meaningful, practical interpretation of the well-known fact that as \( n \to \infty \):

\[ \text{Var} \left[ \frac{X_1 + X_2 + \ldots + X_n}{n} \right] = \frac{\sigma^2}{n} \to 0, \]

where \( X_i, i = 1,2,\ldots \) are i.i.d. random variables with variances \( \sigma \).

### 3. SHARED CONTINUOUS RESOURCE

Consider, firstly, one component that starts operating at \( t = 0 \). Assume that in the process of production (engineering applications) or birth (biological applications) it had acquired an initial unobserved resource \( R \) \([3,4]\). For mechanical or electronic items, for instance, it can be a ‘distance’ between the initial value of the key parameter and the boundary that defines a failure of the component. It is natural to assume that it is a continuous random variable with the Cdf \( F_0(r) \) (the discrete case, as in the previous section, can be considered as well):

\[ F_0(r) = P(R \leq r). \]  

(7)

A battery charged to an unobserved level or the ‘vitality’ of an organism \([11]\) can be considered as relevant examples. A similar notion of a random resource (hazard potential) was considered in references \([2,7]\). Suppose for simplicity that for each realization of \( R \) the component’s remaining resource is monotonically decreasing with time. Therefore, the run out resource, to be called \( \text{wear} \), monotonically increases. The wear in \([0,t)\) can be defined as
\[ W(t) = \int_0^t w(u) du, \]  

where \( w(t) \) denotes the rate of wear. The failure occurs when the wear \( W(t) \) reaches \( R \). Denote the corresponding random time by \( T \). It is clear that

\[ P(T \leq t) = P(R \leq W(t)) = F_0(W(t)). \]  

Therefore, we arrive at a conclusion that this specified survival model can be interpreted in terms of accelerated life model [1]. The generalization to the case of the monotonically increasing random wear process \( W_i, t \geq 0 \) is straightforward [3,5].

Consider now the series system of \( n \) components of the described type with shared resources:

\[ F_i(t) = F_0(W_i(t)); \ i = 1,2,\ldots,n. \]  

Thus, the resources are ‘consumed’ with the ‘accumulated’ rate \( \sum_{i=1}^n w_i(t) \) and the failure of the system occurs when the increasing \( \sum_{i=1}^n W_i(t) \) reaches the accumulated resource \( \sum_{i=1}^n R_i \). Denote the Cdf of \( \sum_{i=1}^n R_i \) by \( F_{\Sigma}(t) \). Then, similar to equation (9), the lifetime Cdf of this system with shared resources is given by

\[ F_s(t) = F_{\Sigma}\left(\sum_{i=1}^n W_i(t)\right). \]  

Equation (11) states the general result for the series system with shared resources. The corresponding survival probability is obviously larger than for the case of individual resources. To proceed further and to obtain relationships similar to (5)-(6), we must first consider the case of identically distributed lifetimes of components with constant in time rates \( w(t) \), i.e., \( F_i(t) = F_0(wt) \). Then, similar to relationship (5), for \( n \to \infty \):

\[ F_s(t) \to \Phi\left(\frac{t - \mu}{\sigma \sqrt{n}}\right) = N\left(\mu, \frac{\sigma^2}{n}\right), \]  

where \( \mu \) and \( \sigma \) are the mean and the standard deviation, defined by the Cdf \( F_0(wt) \).

Indeed, the rate of wear (resource consumption) for the system, which can be considered as some integrated component, is now \( nw \). As the pattern of the overall resource consumption does not matter, let the resource of each original component be consumed consequently with this rate, thus forming the sum of i.i.d. random variables with the Cdf \( F_0(nwt) \) each. Therefore, the probabilistic interpretation is similar to the one of the previous section. In addition, (12) can be easily adjusted to the case of different \( w_i \) because there will be still the sum of i.i.d random variables (except the last cycle).

In the next section we will consider sharing of resources of a different type.
4. OPTIMAL ‘TIME SHARING’ FOR A STANBY SYSTEM WITH AGING COMPONENTS

Consider a standby system of two i.i.d. components. Assume that the failure rate of a component \( \lambda(t) \) is increasing. Therefore, the corresponding Cdf \( F_c(t) \) belongs to the IFR class of distributions. The first component starts operating at \( t = 0 \) and is instantaneously switched to the second one upon its failure. Assume now that switching from the operating component to the standby one can be instantaneously performed at any instant of time. The former operating component then starts to be a standby one and vice versa. We will describe now a simple optimal switching strategy that maximizes the probability of system’s performance without any failures of components in the given time interval. It is clear that the strategy of this kind cannot change the probability of system’s failure (when both components are failed), but an extension of the operation period without failures of components can be very important in various applications (see also the next section).

Denote by \( P_1(S,t) \) the probability of the system’s performance without failures of components in \( (0,t] \) under an arbitrary switching strategy \( S \). Let \( P_1(a,t) \) denote this probability with only one switching at \( a \). Then, it is easy to prove that

\[
\max_S (P_1(S,t)) = P_1\left(a = \frac{t}{2}, t\right) .
\] (13)

Indeed, assuming only one switching, the following minimum should be obtained:

\[
\min_a \left\{ \int_0^a \lambda(u)du + \int_a^t \lambda(u)du \right\}, \ a \in (0,t],
\] (14)

which simply follows from

\[
P_1(a,t) = \exp\left\{ -\int_0^a \lambda(u)du \right\} \exp\left\{ -\int_a^t \lambda(u)du \right\}.
\]

After differentiating the sum of integrals in (14) and equating the result to zero, we arrive at the following equation with respect to an optimal \( a \):

\[
\lambda(a) = \lambda(t-a).
\] (15)

Equation (15) has a trivial unique solution \( a = t/2 \) for increasing functions. It can be also easily seen that, due to additivity of integrals, additional switching cannot improve this result if the total time of functioning of each component is \( t/2 \).

Thus, the described strategy maximizes the probability of system’s performance without failures of components. We will now derive a strategy that maximizes the mean time to the first failure of a component in this system. Denote by \( S_a \) the following strategy. Assume that the first component functions in \( (0,\Delta t] \), then it is switched to the second component, which functions in \( (\Delta t,2\Delta t] \), then again the first component functions in \( (2\Delta t,3\Delta t] \), etc. Denote by \( \lambda_a(t), F_a(t) \) the corresponding
failure rate and the Cdf of time to the first failure. As \( \lambda(t) \) is increasing, the following piecewise constant function is the upper bound for \( \hat{\lambda}_\Delta(t) \):

\[
\hat{\lambda}_\Delta(t) = \begin{cases} 
\lambda(\Delta t), & t \in (0, 2\Delta t] \\
\lambda(2\Delta t), & t \in (2\Delta t, 4\Delta t] \\
\lambda(3\Delta t), & t \in (4\Delta t, 6\Delta t] \\
\vdots & \vdots \\
\lambda(t), & t \in (T, 2T] \\
\vdots & \vdots \\
\lambda(T), & t \in (T, \infty]
\end{cases}
\]

(16)

whereas the corresponding lower bound is

\[
\tilde{\lambda}_\Delta(t) = \begin{cases} 
0, & t \in (0, 2\Delta t] \\
\lambda(\Delta t), & t \in (2\Delta t, 4\Delta t] \\
\lambda(2\Delta t), & t \in (4\Delta t, 6\Delta t] \\
\vdots & \vdots \\
\lambda(t), & t \in (T, 2T] \\
\vdots & \vdots \\
\lambda(T), & t \in (T, \infty]
\end{cases}
\]

(17)

This means that \( \tilde{\lambda}_\Delta(t) \leq \lambda_\Delta(t) \leq \hat{\lambda}_\Delta(t) \) and

\[
\int_0^t \tilde{\lambda}_\Delta(u) du \leq \int_0^t \lambda_\Delta(u) du \leq \int_0^t \hat{\lambda}_\Delta(u) du.
\]

Let now \( \Delta t \to 0 \) and denote the corresponding ‘limiting strategy’ by \( S_i \). It can easily be seen from definitions (16) and (17) that, as \( \Delta t \to 0 \):

\[
\left| \hat{\lambda}_\Delta(t) - \hat{\lambda}\left(\frac{t}{2}\right) \right| \to 0
\]

uniformly in each bounded interval. Finally,

\[
\lambda_i(t) = \hat{\lambda}\left(\frac{t}{2}\right),
\]

(18)

where \( \lambda_i(t) \) denotes the limit failure rate, which has an infinite (countable) number of ordinary points of discontinuity and is integrable (with respect to Lebesgue measure). This, as \( \Delta t \to 0 \), leads to the following weak convergence result:

\[
|F_\Delta(t) - F_i(t)| \to 0
\]

\[
\Rightarrow F_i(t) = 1 - \exp\left\{ -\int_0^t \lambda\left(\frac{u}{2}\right) du \right\}. \tag{19}
\]

It follows from (18) and (19), that the defined operation is a scale transformation of the underlying failure rate \( \lambda(t) \) with a scaling factor \( 1/2 \), and that \( F_i(t) \neq F(t/2) \).

Therefore, by means of strategy \( S_i \), the ‘integrated system’, which shares in the described way the resources of both aging components, had been constructed. It follows from relationship (13) that
\[ P_i(S, t) = P_i\left(\frac{t}{2}, t\right), \quad (20) \]

as the total scheduled time of each component’s operation for both strategies is equal to \( t/2 \). What is the reason for considering the limit strategy instead of the simplest strategy with one switching? It can be easily shown that \( S_l \) maximizes the mean time to the first failure of a component, \( T_l(S) \) and this can be relevant in various applications. Indeed,

\[
\max_S T_l(S) = \max_S \int_0^\infty P_i(S, u) du = \int_0^\infty \exp\left(-\int_0^u \lambda_0 \, du\right) \, du, \quad (21)
\]

as \( S_l \) maximizes \( P_i(S, u) \) for each \( u \in (0, \infty) \). Hence, the random time to the first failure is stochastically larger \([8,9]\) under the strategy \( S_l \) than under any other strategy. Therefore, the corresponding inequality for the means trivially holds.

Along with maximization of the mean time to the first failure of a component, maximization of the time to this event can be useful in many applications. For example, assume that operation of our system can be terminated by some external random event. It can be some other device in series with our system or a shock affecting both components. As \( S_l \) maximizes \( P_i(S, u) \) for each \( u \in (0, \infty) \), this strategy is optimal in the presence of the described random termination.

The reasoning of this section will be 'more pronounced' in the next section, where we consider a system, which quality of performance depends on the number of operating components.

**Remark 2.** The strategy \( S_l \) is, of course, a mathematical idealization. It is obvious that in practice \( \Delta t \) cannot be very small because each preventive switching requires some efforts. The switching device can also be unreliable, but, unlike the switching upon failure, the state of this device can be checked prior to the preventive switching. Thus, in practice \( S_l \) means that switching should be performed as often as reasonable. Given the corresponding costs and rewards, a problem of obtaining some optimal \( \Delta t_{op} \) can be considered.

**Remark 3.** The considered approach can be generalized in a straightforward way to a standby system of \( n > 2 \) i.i.d. components. The strategy, maximizing the time to the first failure, is the one when \( n-1 \) equidistant switching are scheduled. Then, using the same approach, \( P_2(S, t_1, t) \) - the probability of system of \( n-1 \) components performance without failures in \((t_1, t]\), where \( t_1 \) is the time of the first failure, can be maximized, etc. The corresponding limiting strategy is also defined in an obvious way. Another generalization is to the case of the non-identical components with increasing failure rates \( \lambda_1(t) \) and \( \lambda_2(t) \). The optimal time of switching \( a \) can be uniquely obtained from the equation similar to (15):

\[ \lambda_1(t) = \lambda_2(t - a). \]
The corresponding limiting strategy also can be defined.

**Example 2.** Let \( F_c(t) \) be exponential, i.e., \( F_c(t) = 1 - \exp\{-\lambda t\} \). Assume that each component of this system can be instantly perfectly repaired upon failure, but the number of repairs is bounded by \( m \geq 0 \), as, for instance, in the case of spare parts. The ‘total failure’ of each component occurs when \( m \) repairs had been already performed and a component fails. Therefore, the Cdf of time to this failure is the Erlangian distribution (2) with the increasing failure rate [6].

What strategy should be used for maximizing the probability of system’s functioning without total failures of components in the given interval of time. The formal answer, based on the previous results, is that the only switching at \( t = a/2 \) (and the \( S_f \) strategy for maximizing \( T_i \)) should be used for this purpose. On the other hand, now we are able to perform the dynamic strategy of switching, based on the additional information. This information is the number of repairs left for each component at any instant of time. From a simple probabilistic reasoning it follows that an optimal strategy in this case is any strategy \( S_{opt} \) that leaves the remaining component (after the total failure of the other one) in the state with no repairs left. Therefore,

\[
P_f(S_{opt}, t) = \exp\{-\lambda t\} \sum_{i=0}^{m} \frac{(\lambda t)^i}{i!}.
\]

For any other strategy, including the one with switching at \( a = t/2 \) and \( S_f \), we have

\[
P_f(S_{opt}, t) \geq P_f(S, t).
\]

Therefore, additional information can increase the ‘quality’ of the switching policy.

**5. OPTIMAL STRATEGIES FOR FINITE INTERVALS OF TIME**

Consider now the operation of the standby system (described in Section 4) to its total failure when the second component also fails. It is obvious that instantaneous, perfect switching from one component to another cannot change the corresponding survival function, but it can increase the quality of performance in the finite interval of time. Therefore, assume additionally that each operating component is characterized by the reward function \( w_h(x) \), i.e., the reward in \((x, x + dx)\) is \( w_h(x)dx + o(dx)\) and the accumulated reward (profit) in \([0, t]\) (without failures) is

\[
W_h(t) = \int_0^t w_h(x)dx.
\]

Let \( w_h(x) \) be a decreasing function, which describes deterioration (ageing) with time. The reward in the failed state is assumed to be 0. Therefore, the expectation of the reward accumulated by one component in \([0, t]\) is

\[
R(0, t) = \overline{F}_c(t)W_h(t) + \int_0^t f_c(x)W_h(x)dx,
\]

(22)
where \( f_c(x) = F'_c(x) \) and \( F_c(t) = 1 - F_c(t) \).

Similar to the previous section, we want to obtain a strategy of switching that will maximize the system’s reward \( R_s(0,t) \). It is clear that \( R_s(0,\infty) \) does not depend on switching and is equal to \( 2R(0,\infty) \), i.e.,

\[
R_s(0,\infty) = 2\int_0^\infty f_c(x)W_R(x)dx, \quad (23)
\]

where we assume, for simplicity, that \( \lim_{t \to \infty} F_c(t)W_R(t) = 0 \), which given that \( W_R(t) \) cannot increase faster than a linear function, is a very mild condition on the Cdf of a component \( F_c(t) \). Note that, the heavy-tailed Pareto distribution for certain values of parameters does not comply with this condition.

However, for the finite interval, switching can make sense as it ‘moves’ intervals of time with smaller values of reward outside the mission interval \([0,t]\). Let a single switching be performed at \( a \). The corresponding reward in \([0,t]\) is given by the following sum:

\[
R_s(0,t,a) = F_c(a)F_c(t-a)(W_R(a) + W_R(t-a)) \\
+ \int_0^a f_c(x)F_c(t-x)(W_R(x) + W_R(t-x))dx \\
+ \int_a^t f_c(x-a)F_c(a+t-x)(W_R(a+t-x) + W_R(x-a)) \\
+ \int_{x=2<t} f_c(x)W_R(x) + W_R(y)dy.
\]

The last term, which corresponds to realizations, where both components had failed in \([0,t]\), similar to (23), does not depend on switching. Differentiation of \( R_s(0,t,a) \) with respect to \( a \) results in the following equation:

\[
R'_s(0,t,a) = F_c(a)F_c(t-a)(W(a) - W(t-a)).
\]

Therefore, as \( W_R(t) \) is a monotonically increasing function,

\[
R'_s(0,t,a) = 0 \Rightarrow W_R(a) = W_R(t-a) \Rightarrow a = \frac{t}{2},
\]

and it can be proved that \( R'_s(0,t,a) \) achieves maximum at \( a = t/2 \).

It can be also proved that additional switching does not improve this result. Therefore, switching (if any) should be planned at \( a = t/2 \). The limiting strategy of the previous section is considered in a similar way and the case of components with different decreasing reward functions \( W_{R_i}, i=1,2 \) as well. Specifically, the optimal point for the latter case is uniquely derived from the equation

\[
W_{R1}(a) = W_{R2}(t-a) = 0.
\]
**Example 3** (explanatory). Assume that the components lifetimes are described by the degenerate distributions with a mass at $b$ and exponential decay function $\exp\{-\alpha t\}$. Let $t < b$ (the similar considerations hold for $b < t < 2b$, whereas there is no need in switching when $t > 2b$, as the switching is ‘automatically’ performed at $t = b$ upon failure).

The reward without switching is just the reward from the first component, \textit{i.e.},

$$R(0,t) = \int_0^t \exp\{-\alpha u\} du = \frac{1}{\alpha} (1 - \exp\{-\alpha t\}),$$

(24)

whereas the reward with switching at $a = t/2$ is

$$R_s(0,t, t/2) = 2 \int_0^{t/2} \exp\{-\alpha t\} dt = \frac{2}{\alpha} \left(1 - \exp\{-\alpha (t/2)\}\right),$$

which is larger than $R(0,t)$.

**6. CONCLUDING REMARKS**

In this paper, we have considered different types of resource sharing. It is clear from general considerations that sharing of resources of different parts of a system can increase its performance quality. For example, sharing of spare parts increases probability of spare parts sufficiency for a system in the mission period $[0,t]$. It is shown that when a large number of identical components in series share resources, the failure time distribution for a series system tends to a degenerate distribution. This means that the corresponding reliability of a system under certain assumptions can be close to 1 interval $[0,m]$, where $m$ is the mean lifetime of a component, and then falls sharply to 0 outside this interval.

Another type of sharing resources arises when considering performance of standby systems with aging components. We understand aging as increasing of the corresponding failure rate or as decreasing of a performance characteristic of a component and derive simple optimal switching strategies. Specifically, we show that the switching at $a = t/2$, maximizes the probability of the system’s performance without failures of components in $[0,t]$ and the switching strategy which performs switching at $n\Delta t, n = 1,2,...$, where $\Delta t \to 0$, maximizes the mean time to this event.

**References**


