Simulation Study for the Optimal Repair Capacity of an IT Maintenance Center based on LRD Failure Distribution

So Young Sohn
Dept of Computer Science & Industrial
Systems Engineering, Yonsei University
134 Shinchondong Sodemungu
Seoul Korea
sohns@yonsei.ac.kr

Ki Yeol Oh
Dept of Computer Science & Industrial
Systems Engineering, Yonsei University
ky6450@hanmail.net

Abstract

Consider various host systems that consist of both hardware and software. There is an IT center which governs the activities of all of those host systems including the reported failures. The center's major concern is deciding the proper level of repair capacity for various kinds of failures due to hardware, software, and operator related problems where corresponding failure and repair times may exhibit fairly large variations. In this paper, we use Monte Carlo simulation in order to decide the optimal repair capacity when both failures and repair times of individual host systems exhibit long range dependent (LRD) property. In terms of both utilization rate and the maximum waiting time, we suggest optimal capacity. Various sensitivity analyses follow.

Keywords: long range dependent property, system failure, multi response optimization

1. Introduction

Allocation of optimal resources for software/ hardware maintenance is an important topic in the era of Information age. In the typical Information software/ hardware maintenance problem, classical lifetime distributions such as exponential, Weibull and lognormal have been used to model both the failures and repair times.

Subsequently, queueing models are formed according to both failure and repair distributions and the number of servers for repair. Traditional queueing models, and many simulation models of queues, assume that such failure/ service times are independent and identically distributed. However, it is now well known that some queueing systems exhibit long-range dependence property. Long-range dependence (LRD) property is characterized by very slowly decaying autocorrelations and by sample means whose variance decreases with sample size n at a rate slower than 1/n (Beran et al.1995).

In this paper, we apply LRD property to input distributions for failure/repair times of software/hardware. The main purpose is to find the optimal resources for IT system maintenance in consideration of the LRD failure/repair distribution. We perform various sensitivity analyses due to changes in both performance measures and identification of input distributions.

2. Long Range Dependence

It is interesting to note that the Pareto distribution can be derived from a random effects model context where the arrival/service times are assumed to follow exponential distributions for given rates where they are the outcomes of gamma distribution.

Suppose that for given arrival rate λ_{ik} and service rate μ_{ik}, the i-th inter-arrival time (U_{ik}, i=1,...,n) and the j-th service time (V_{jk}, j=1,…,m) for the failure reported from system k follows exponential distributions:

$$f(u_{ik} \mid \lambda_{ik}) = \lambda_{ik} \exp(-\lambda_{ik} u_{ik}) \cdots\cdots (1)$$

$$f(v_{ik} \mid \mu_{ik}) = \mu_{ik} \exp(-\mu_{ik} v_{ik})$$

We assume that the arrival rate λ_{ik} and service rate μ_{ik} would vary due in part to the surrounding conditions of each system, represented by 1 x p covariate vectors, x_{ik} and z_{ik} , respectively. The random variations in the λ_{ik} and μ_{ik} that cannot be explained by these conditions are assumed to be related to a Gamma random variable T with density $g(t;a,b) = b^a / \Gamma(a)\, t^{a-1} \exp(-bt)$. That is T ~ Gamma(a,b).

Consequently,

$$\lambda_{ik} = \exp(x_{ik}\gamma + \varepsilon_{ik}) = \exp(x_{ik}\gamma) e_{ik}$$

$$\mu_{ik} = \exp(z_{ik}\theta + \delta_{ik}) = \exp(z_{ik}\theta) d_{ik} \cdots\cdots (2)$$

where e_{ik} ~ Gamma (α+1 , α), d_{ik} ~ Gamma (β+1 , β). The reason of specific parameterization in Gamma distributions for both e_{ik} and d_{ik} is to ensure the expected values of both arrival and service to be $1/\exp(x_{ik}\gamma)$ and $1/\exp(z_{ik}\theta)$, respectively. Use of a gamma prior distribution in (2) may be considered subjective. But the choice of associated parameters such as (α,β,γ,θ) can give ample flexibility of modeling random patterns.

Based on (1) and (2), the joint distribution of (α,β,γ,θ) can be obtained as follows:

$$f(u_{ik}, v_{ik}) = \int f(u_{ik} \mid \lambda_{ik}) g(\lambda_{ik} : \alpha, \alpha / \exp(x_{ik}\gamma)) d\lambda_{ik} \int f(v_{ik} \mid \mu_{ik}) g(\mu_{ik} : \beta, \beta / \exp(z_{ik}\theta)) d\mu_{ik}$$

$$= \frac{(\alpha+1)\alpha^{\alpha+1} \exp(x_{ik}\gamma)}{(\alpha + \exp(x_{ik}\gamma) u_{ik})^{2+\alpha}} \frac{(\beta+1)\beta^{\beta+1} \exp(z_{ik}\theta)}{(\beta + \exp(z_{ik}\theta) v_{ik})^{2+\beta}} \cdots\cdots (3)$$

It shows that the probability density functions of u_{ik} and v_{ik} are sometimes called Lomax distribution or Pareto type VI distribution, where $E(u_{ik})=E(E(u_{ik} \mid e_{ik}))=E(1/\lambda_{ik})$ and $Var(u_{ik}) = Var(E(u_{ik} \mid e_{ik})) + E(Var(u_{ik} \mid e_{ik})) = Var(1/\lambda_{ik}) + E(1/\lambda_{ik}^2)$. Note that $1/\lambda_{ik}$ ~ Inverse gamma (α+1, α/exp($x_{ik}\gamma$)) with mean 1/exp($x_{ik}\gamma$) and variance $1/[\exp(x_{ik})]^2 (\alpha-1)$. Therefore both the expected arrival time and variance can be derived as follows:

$E(u_{ik})=1/\exp(x_{ik}\gamma)$ ………………………………………………………………… (4)

$Var(u_{ik})=2/[\exp(x_{ik}\gamma)]^2(\alpha-1) + 1/[\exp(x_{ik}\gamma)]^2$

$=(\alpha+1)/[\exp(x_{ik}\gamma)]^2(\alpha-1)$, if $\alpha>1$ ………………………………………… (5)

Therefore according to (5), Pareto with covariates becomes LRD, when α is less than 1.

For the simulation of Pareto random variables u_{ik} and $v_{ik,}$ the following cumulative distribution for u_{ik} and v_{ik} can be applied:

$$F(u_{ik}) = \frac{\alpha^{\alpha+1}\exp(x_{ik}\gamma)}{(\alpha+\exp(x_{ik}\gamma)u_{ik})^{1+\alpha}}$$

$$F(v_{ik}) = \frac{\beta^{\beta+1}\exp(z_{ik}\theta)}{(\beta+\exp(z_{ik}\theta)v_{ik})^{1+\beta}} \quad \cdots\cdots (6)$$

F(u) and F(v) in (6) follow uniform distributions and one can generate Pareto random variables u_{ik} and v_{ik} as follows:

$$u_{ik} = \frac{\alpha}{\exp(x_k\gamma)}\left[\frac{1}{(1-F(u_{ik}))^{1/(1+\alpha)}}-1\right]$$

$$v_{ik} = \frac{\beta}{\exp(z_k\theta)}\left[\frac{1}{(1-F(v_{ik}))^{1/(1+\beta)}}-1\right] \quad \cdots\cdots (7)$$

In particular, LRD Pareto distribution can be simulated by adjusting the ranges of α and β to be less than 1(Sohn 2002).

3. Design of Experiment

We first describe a motivating case. Consider various host systems that consist of both hardware and software for seven branches.

There is a center that governs the activities of all of those host systems including the repair of reported failures. The center's major concern is assessing the proper level of repair capacity for those failures. Failures might be due to hardware, software, or operator related problems. In this paper, we use Monte Carlo simulation based experimentation in order to decide the optimal repair capacity when both failures and repair times of individual host systems exhibit long range dependent (LRD) property.

For simulation design, we assume the following, based on the observed data in IT center of S group in Korea:

1. A total number of host systems is seven (k=1,...,7).
2. Sources of failures are three: hardware, software, and operator. Composition of these failures is 25%, 54% and 21% for hardware, software and operator, respectively.

3. The i-th inter-failure time reported by host system k, u_{ik}, follows Pareto distribution with $E(u_{ik})=1/\exp(x_{ik}\gamma)$ and $Var(u_{ik})= E(u_{ik})^2(\alpha +1)/(\alpha -1)$. Based on the previous year's experience at S group, we assume the following:

$E(u_{ik})$= 72, 51.4, 24, 8.2, 36, 7.3, 22.5 days for the seven companies respectively.

Accordingly, x_{ik} in (7) is a 1 x 7 vector distinguishing the seven company effects: we consider the company with as $E(u_{ik})$= 72 as a reference group and set intercept as –7.455 so that the mean arrival time to be 72days x 24 hours=1/exp(-7.455). The remaining γ values in (7**)** are set 0.624, 1.099, 1.511, 0.756, 2.08 and 1.163, respectively. To set the variance of x_{ik}, values for α need to be assigned.

In view of the fact that LRD appears when $0 < \alpha < 1$, we consider α to be the first factor with two levels: 0.5 and 1.5, where 0.5 is used of LRD and 1.5 is for non-LRD.

4. The j-th service time for the failure reported from system k, v_{jk}, follows Pareto distribution with $E(v_{jk})=1/\exp(z_{ik}\Theta)$ and $Var(v_{jk})= E(v_{jk})^2(\beta +1)/(\beta -1)$. We assume that $E(v_{jk})$= 4 hours, if the j-th service was for hardware or software failure; and it is 2 hours, if the j-th service was for operator related failures. This assumption also reflects the previous experience in S group in Korea. Again we set operator as a reference group and let $\exp(z_{ik}\Theta) =\exp(-0.693-0.693z1_{ik})$ where z1 is indicator variables representing additional effects of hardware and software to operator respectively. Next, we consider β to be the second factor with two levels (0.5 and 1.5) to represent LRD and non-LRD service distribution.

All these settings are based on the current repair system used in S group. In order to consider the changes in repair capacity for hardware, software and operator related failures, we consider them as additional factors to α and β. Note that the two factors α and β are typically uncontrollable while the remaining three are controllable.

Acknowledgement

This work was supported by Korea Research Foundation Grant (KRF-2001-041-E00119).

References

Beran J, Sherman R, Taqqu MS, Willinger W.(1995). Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on communications. 1995*;43:1566-79.

.

Sohn, So Young (2002), "Robust design of server capability in M/M/1 queues with both partly random arrival and service rates", *Computers & Operations Research* 29, pp.433-440