255/2010

Raport Badawczy

Research Report

RB/9/2010

Improving accuracy of reliability parameters estimation for a commodity flow network by means of stratified sampling

J. Malinowski

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2010

1. Introduction

A commodity transportation network composed of nodes and links arranged in a tree structure is considered. Its task is to transfer a commodity (electric power, radio signal, electronic data, water, gas, etc.) from the source (root) node to all terminal (leaf) nodes. Let $\{e_0, e_1, ..., e_m\}$ be the set of all network components, i.e. its nodes and links. The components are indexed so that the index of each non-root component is greater than the index of its parent component, e_0 being the root component, i.e. the root node. An exemplary network consisting of 13 nodes and 12 links is presented in Fig. 1.

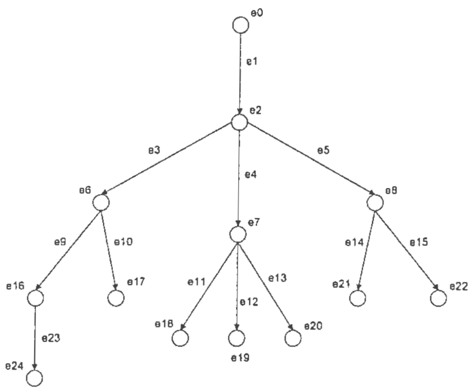

Figure 1. An exemplary system structure

Each component can be in one of two states: operable – 1 and failed – 0; e_0 is always in the operable state. The repair of a failed component begins as soon as one of repair teams is available – due to their limited number a waiting period following a failure may occur. The order in which failed components are chosen for repair depends on the repair policy applied – two types of policies will be considered. The functioning of the component e_i is characterized by three distribution functions: F_i – lifetime d.f. of the operable e_i connected to e_0, G_i – lifetime d.f. of the operable e_i disconnected from e_0, and H_i – repair time d.f. of the failed e_i. **It is assumed that F_i and G_i are exponential**, unlike H_i which can be arbitrary d.f. on $[0,\infty)$. It is also assumed that $F_i \geq G_i$, which conveys the idea that the components being "under load" are more failure prone. Thus, a component's lifetime depends on the behavior of all „upstream" components but is not influenced by the remaining components. However, a component functions independently of the "upstream" components up to the moment when one of them fails. Furthermore, a component's repair time is independent of the states of all other components. Note that $G_i \equiv 0$ if it is assumed that e_i disconnected from e_0 cannot fail.

A commodity can be transferred from e_0 to e_i, $i = 1,\ldots, m$, if and only if e_i is functional and connected to e_0, i.e. all components between e_0 and e_i are in the operable state. As failures of components occur, the periods of connection between e_0 and the operable e_i are interlaced by the periods during which e_i is failed or disconnected from e_0. Our objective is to determine the mean durations of both time intervals, i.e. the equivalents of MTBF and MTTR parameters, for each component. As statistical estimation will be used, special attention will be paid to its accuracy.

Clearly, the durations of connection and disconnection periods depend on two more factors: the number of repair teams assigned to the network maintenance, and the repair policy implemented. Obviously, each component's average disconnection time decreases as the number of repair teams increases, because the average time a component waits for a repair to commence becomes shorter. As to the second factor, three repair policies will be considered. According to the first policy the components are chosen for repair in the order in which they failed, i.e. they form a FIFO queue; if multiple components fail at the same time (**such event occurs with zero probability unless it is a common cause failure**), the one with the largest index is selected as first. This policy will be named "FIFO with largest index priority". If $G_i \equiv 0$, $1 \leq i \leq m$, then for each "linear" subset of components (i.e. all components located between e_0 and a leaf node) it prioritizes the components most distant from e_0. Indeed, as only the components connected to e_0 can fail, if e_y

is located below e_x (yielding y>x), then e_y can only fail if e_x is operable, i.e. e_y can only fail before or simultaneously with e_x, hence e_y must precede e_x in the queue for repair.

According to the second policy the components are selected for repair in the order reverse to that in which they failed, i.e they form a LIFO queue; if multiple components fail at the same time, the one with the smallest index is selected as first. This policy will be named "LIFO with smallest index priority". If $G_i \equiv 0$, $1 \leq i \leq m$, then for each "linear" subset of components it prioritizes the components least distant from e_0. Indeed, if e_y is located above e_x (yielding y<x), then e_y can only fail after or simultaneously with e_x, hence e_y must precede e_x in the queue for repair.

The third policy prioritizes the components according to their indexes, i.e. the first component in the queue for repair is the one with the smallest index. This policy will be named "smallest index priority". Note that it does not take into account the order in which the components fail, and is only determined by the numbering of the components. Certainly, in general case, the numbering scheme reflecting a particular repair policy for a multi-component system can be different from the one adopted in this paper.

The described system has been first investigated in [Malinowski 2009], where the simulation technique used to imitate the system's failure-repair process has been presented in detail, along with some theoretical results concerning that process. The sought parameters have been evaluated using interval estimation, and the non-trivial problem of finding their confidence limits has been tackled. This paper focuses on improving the accuracy with which the considered reliability parameters are estimated. In order to achieve that aim the appropriately modified "stratified sampling" method – one of the so-called variance reduction techniques – is used. The above mentioned modification (devised by the author of this paper) is necessary due to the fact that the basic estimated parameter is the quotient of mean values of two dependent random variables rather than a single mean value.

2. Definitions and Notation

The following notation will be used in the paper:

$L_i^{(1)}$, $L_i^{(2)}$ – lifetimes of operable e_i connected to/disconnected from e_0

R_i – repair time of e_i

r_{min}, r_{max} – minimum and maximum repair time of a single component

F_i, G_i, H_i – distribution functions of $L_i^{(1)}$, $L_i^{(2)}$, and R_i

λ_i – failure intensity of operable e_i connected to e_0

π_1 – the "FIFO with largest index priority" policy

π_2 – the "LIFO with smallest index priority" policy

π_3 – the "smallest index priority" policy

To avoid over-indexing, it is assumed in the remaining definitions that there are r repair teams and the policy π_s is applied.

$A_j^{(i)}$ – length of the j-th period during which operable e_i remains connected to e_0 (independent of r and s under the above assumptions), $j \geq 1$

$B_j^{(i)}$ – length of the j-th period during which e_i remains failed or disconnected from e_0, $j \geq 1$

$a^{(i)}$, $b^{(i)}$ – the average durations of $A_j^{(i)}$ and $B_j^{(i)}$ over $j \geq 1$

$c^{(i)}$ – the average number of restored connections between operable e_i and e_0 per unit time

X – the system's failure-repair process, i.e. the vector-valued random process $\{[X_1(t),...,X_m(t)], t \geq 0\}$, where $X_i(t)$ is defined as follows:

$X_i^{(t)} = -q$, if e_i is in place q in the repair queue

$X_i^{(t)} = 0$, if e_i is under repair

$X_i^{(t)} = 1$, if e_i is operable and connected to e_0

$Xi^{(t)} = 2$, if e_i is operable and disconnected from e_0

τ_k – the moment of k-th recurrence of X to its initial state, i.e. the state (1,…,1), $k\geq 0$, $\tau_0=0$

$Q^{(i)}$ – the number of restored connections between operable e_i and e_0 in the interval $(\tau_0, \tau_1]$

$U^{(i)}$ – the total time within $(\tau_0, \tau_1]$ during which e_i remains failed or disconnected from e_0

We thus have:

$$a^{(i)} = \lim_{m\to\infty} \frac{1}{n}\sum_{j=1}^{n} A_j(i), \quad b^{(i)} = \lim_{m\to\infty} \frac{1}{n}\sum_{j=1}^{n} B_j(i) \tag{1}$$

It has been shown in [Malinowski 2009] that

$$a^{(i)} = \frac{1}{\sum(\lambda_j : e_j \lhd e_i)} \tag{2}$$

where $e_j \lhd e_i$ means that e_j is located between e_0 and e_i. On the other hand, $b^{(i)}$ has to be estimated (provided it is a constant value) as in the general case finding a close formula for $b^{(i)}$ is practically impossible.

Each interval $[\tau_{k-1}, \tau_k)$, $k\geq 1$, will be called an operational cycle of X. As $L_1^{(1)}, \ldots, L_m^{(1)}$ are exponentially distributed and independent, X can be divided into the independent and stochastically identical sub-processes $\{X(t), t\in[\tau_{k-1}, \tau_k)\}$, $k\geq 1$. This explains why $Q^{(i)}$ and $U^{(i)}$ are only defined for the first cycle – for all subsequent cycles they are identically distributed.

3. Estimating $b^{(i)}$ with the use of Monte Carlo simulation

In [Malinowski 2009] the algorithm for the simulation of X is presented. Also, several lemmas constituting theoretical basis for the estimation of $b^{(i)}$ are given. One of them formulates certain necessary conditions under which X is a recurrent process. Another lemma states that if X is recurrent, and repair times are bounded from above and below, i.e. $0 < r_{min} \leq R_i \leq r_{max} < \infty$ for $1 \leq i \leq m$, then $[B_1^{(i)}+\ldots+B_n^{(i)}]/n$ converges in probability to $E[U^{(i)}]/E[Q^{(i)}]$ as $n \to \infty$, then

$$b^{(i)} = \frac{E(U^{(i)})}{E(Q^{(i)})} \tag{3}$$

i.e. $b^{(i)}$ is a constant value.

Let us note that a process $\Pi = \{\Pi(t), t \geq 0\}$ is called recurrent if it has the following properties:

1. The state of Π at $t=0$ is fixed, i.e. $\Pi(0)$ has the one-point distribution,
2. With probability one Π returns to the state $\Pi(0)$ after finite time, i.e. $Pr(\tau_1<\infty)=1$, where τ_1 is the (random) time of the first return of Π to its initial state
3. $\Pi_1 = \{\Pi(\tau_1+t), t\geq 0\}$ and Π are stochastically identical processes (Π "begins anew" at $t = \tau_1$).

Sometimes the second property is replaced with a stronger one, i.e. $E(\tau_1)<\infty$. For details see [Feller 1968].

According to (3), the quotient of the estimates of $E(U^{(i)})$ and $E(Q^{(i)})$ will be used as the estimate of $b^{(i)}$. A natural question arises – what is the accuracy of such estimate? In terms of interval estimation this problem consists in finding a confidence interval for the quotient of two non-independent random variables. The sought confidence interval is given by the following lemma, the proof of which is given in [Malinowski 2009].

Lemma 1

Let $X \geq 0$ and $Y \geq y_{min} > 0$ be random variables with finite means μ_X and μ_Y, and finite standard deviations σ_X and σ_Y (y_{min} is a constant). Let $\hat{X}_K$ and $\hat{Y}_L$ be sample means from X and Y of sizes K and L. Let

$$\varepsilon_{\alpha} = \frac{2q_{1-\alpha/4}}{y_{min}^2} \max\left[\frac{\sigma_X \mu_Y}{\sqrt{K}}, \frac{\sigma_Y \mu_X}{\sqrt{L}}\right] \tag{4}$$

where $q_{1-\alpha/4}$ is the $1 - \alpha/4$ quantile of the standardized normal distribution, i.e.

$$\Pr(Z \leq q_{1-\alpha/4}) = 1 - \frac{\alpha}{4} \tag{5}$$

for normally distributed Z with the expected value 0 and variance 1. Then, for sufficiently large K and L we have:

$$\Pr\left(\left|\frac{\hat{X}_K}{\hat{Y}_L} - \frac{\mu_X}{\mu_Y}\right| > \varepsilon_{\alpha}\right) \leq \alpha \tag{6}$$

i.e. $[\hat{X}_K/\hat{Y}_L - \varepsilon_{\alpha},\ \hat{X}_K/\hat{Y}_L + \varepsilon_{\alpha}]$ is a $1 - \alpha$ confidence interval for μ_X/μ_Y.

To further simplify the notation, the symbols U and Q will be used in place of $U^{(i)}$, and $Q^{(i)}$, it being understood that i is the default component's index. Let U* and Q* be conditional random variables defined as follows: $U^* = U|Q \geq 1$ and $Q^* = Q|Q \geq 1$, i.e. the underlying condition is that e_i

fails or is disconnected from e_0 at least once during the interval (τ_{k-1}, τ_k). Thus U* and Q* are defined on $\{\omega: Q(\omega)\geq 1\}$, and their distribution functions $F_{U^*}(u)$ and $F_{Q^*}(q)$ are given by $\Pr(U < u, Q \geq 1)/\Pr(Q \geq 1)$ and $\Pr(Q < q, Q \geq 1)/\Pr(Q \geq 1)$ respectively. As $E(U|Q=0) = 0$, we have:

$$\frac{EU}{EQ} = \frac{E(U,Q\geq 1)+E(U,Q=0)}{E(Q,Q\geq 1)+E(Q,Q=0)} = \frac{E(U\mid Q\geq 1)}{E(Q\mid Q\geq 1)} = \frac{E(U^*)}{E(Q^*)} \tag{7}$$

Thus we can estimate E(U*)/E(Q*) instead of E(U)/E(Q), obtaining the same result.

Note that Lemma 1 makes no assumption about the independence of X and Y, therefore it covers the case of strongly dependent random variables such as U* and Q*. Also note that $Q^*\geq 1>0$, i.e. Q* is bounded from below by a positive constant, which is not true in case of Q. Thus Lemma 1 can be applied to U* and Q*, but not to U and Q. Replacing in Lemma 1 X and Y by U* and Q* we obtain the following corollary: If

$$\varepsilon_\alpha = 2q_{1-\alpha/4}\max\left(\frac{\sigma_{U^*}\mu_{Q^*}}{\sqrt{K}}, \frac{\sigma_{Q^*}\mu_{U^*}}{\sqrt{L}}\right) \tag{8}$$

then, for sufficiently large K and L, $[\, U\hat{}^*_K/Q\hat{}^*_L - \varepsilon_\alpha , U\hat{}^*_K/Q\hat{}^*_L + \varepsilon_\alpha \,]$ is a $1-\alpha$ confidence interval for E(U*)/E(Q*). In consequence, if ε is a given small number and

$$K \geq \left(\frac{2q_{1-\alpha/4}\sigma_{U^*}\mu_{Q^*}}{\varepsilon}\right)^2 \wedge L \geq \left(\frac{2q_{1-\alpha/4}\sigma_{Q^*}\mu_{U^*}}{\varepsilon}\right)^2 \tag{9}$$

then taking at least K samples from U* and L samples from Q* allows to estimate E(U*)/E(Q*) with the accuracy at least ε (the half-length of the confidence interval) at the given confidence level $1-\alpha$. In practice, the expected values and standard deviations of U* and Q* are replaced in (9) with the respective sample means or sample standard deviations.

The above considerations lead to the following algorithm estimating E(U*)/E(Q*).

Algorithm 1

- Perform N_p "pilot simulations" (i.e. simulate N_p operational cycles of *X*) to find approximate values of μ_{U^*}, σ_{U^*}, μ_{Q^*}, and σ_{Q^*}
- Compute minimal values of K and L fulfilling (9)
- Execute the main estimation procedure outlined below

Procedure 1

```
j1 = 0 ; j2 = 0 ; μU* = 0 ; μQ* = 0 ;
do {
   simulate one full operational cycle of X ;
   if (Q≥1) {
      if (j1 < K) {
         j1 = j1 + 1 ; μU* = μU* + (U – μU*) / j1 ;
      }
      if (j2 < L) {
         j2 = j2 + 1 ; μQ* = μQ* + (Q – μQ*) / j2 ;
      }
   }
} while (j1 < K  OR  j2 < L);
return (μU* / μQ* );
```

Remark: μ_{U^*} and μ_{Q^*} are updated based on the following formula:

$$\mu_j = \mu_{j-1} + (x_j - \mu_{j-1})/j \qquad (10)$$

where

$$\mu_j = (x_1 + \ldots + x_j)/j \tag{11}$$

4. Stratified sampling as a means of reducing the computational complexity

Stratified sampling is an estimation technique (one of the so-called variance reduction techniques) used when there are two **dependent** random variables – X with unknown mean value (to be estimated), and a finite-valued random variable Z which fulfills the following condition:

$$\Pr(Z = z) = p_z, \quad 1 \le z \le m, \quad \sum_{z=1}^{m} p_z = 1 \tag{12}$$

Let $X_1,\ldots,X_n$ be a random sample of size n from X, and $X_1^{(z)},\ldots,X_{n_z}^{(z)}$ – a random sample of size n_z from the conditional r. v. X|Z=z defined by its distribution function $F_{X|Z=z}$ in the following way:

$$F_{X|Z=z}(x) = \Pr(X \le x \mid Z = z) = \Pr(X \le x, Z = z)/\Pr(Z = z) \tag{13}$$

It is assumed that the sizes of the random samples from X|Z=z, $1 \le z \le m$, add up to n, i.e. the equality

$$\sum_{z=1}^{m} n_z = n \tag{14}$$

holds. In order to estimate $\mu_X = E(X)$, the "stratified" estimator $\hat{X}_{n,st}$ given by

$$\hat{X}_{n,st} = \sum_{z=1}^{m} p_z \frac{1}{n_z} \sum_{j=1}^{n_z} X_j^{(z)} \tag{15}$$

is used instead of the usual (non-stratified) estimator $\hat{X}_n$, where

$$\hat{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j \tag{16}$$

Both $\hat{X}_n$ and $\hat{X}_{n,st}$ are unbiased estimators, hence the latter can also be used to approximate μ_X. Furthermore, as shown in Glasserman 2003, for large n we have:

$$\Pr(|\hat{X}_{n,st} - \mu_X| \le q_{1-\alpha/2} \frac{\sigma_{X,st}}{\sqrt{n}}) \ge 1 - \alpha \tag{17}$$

where $q_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standardized normal distribution (see (5)), $\sigma^2_{X,st}$ is the so-called stratified variance of X defined as follows:

$$\sigma^2_{X,st} = \sum_{z=1}^{m} \frac{n}{n_z} p_z^2 \sigma^2_{X|Z=z} = n\mathrm{Var}(\hat{X}_{st}) \tag{18}$$

and $\sigma_{X|Z=z}$ is the standard deviation of the conditional r.v. X|Z=z, i.e.

$$\sigma^2_{X|Z=z} = \mathrm{Var}(X \mid Z = z), \quad 1 \le z \le m \tag{19}$$

The second equality in (18) follows from the independence of $X_j^{(z)}$, $1 \le z \le m$, $1 \le j \le n_z$. Thus $\hat{X}_{st}$ and $\sigma_{X,st}$ can be used to compute confidence bounds for μ_X. In practice, $\sigma_{X,st}$ is often unknown, hence it is replaced in (17) by $\hat{\sigma}_{X,st}$ defined as follows:

$$\hat{\sigma}_{X,st} = \sum_{z=1}^{m} \frac{n}{n_z} p_z^2 \hat{\sigma}^2_{X|Z=z} \tag{20}$$

where $\hat{\sigma}^2_{X|Z=z}$ is the sample variance of X|Z=z obtained from its n_z sample values.

Using the well-known fact that

$$\mathrm{Var}(X) = E[\mathrm{Var}(X \mid Z)] + \mathrm{Var}[E(X \mid Z)] \tag{21}$$

where X and Z are arbitrary random variables defined on the same probability space, we obtain:

$$\begin{aligned} \sigma_X^2 &= \mathrm{Var}(X) \ge E[\mathrm{Var}(X \mid Z)] = \\ &= \sum_{z=1}^{m} \mathrm{Var}(X \mid Z = z)\Pr(Z = z) = \sum_{z=1}^{m} p_z \sigma^2_{X|Z=z} \end{aligned} \tag{22}$$

From (22) it follows that if

$$n_z = np_z = n_z^{prop} \tag{23}$$

i.e. the so-called proportional sampling is performed, then

$$\sigma^2_{X,st} = \sum_{z=1}^{m} p_z \sigma^2_{X|Z=z} \le \sigma^2_X \tag{24}$$

Thus, estimating μ_X with the use of $\hat{X}_{st}$ and $\sigma_{X,st}$, we obtain, according to (17), the confidence interval which is narrower in comparison with the one obtained using $\hat{X}$ and σ_X. In other words, proportional sampling allows either to estimate μ_X with greater accuracy without increasing n – the total size of the random sample, or to reduce n while retaining the same accuracy. Applying Lagrange multipliers it can be shown that $\sigma_{X,st}$ attains its minimum value (with respect to n_z, $1 \le z \le m$) if

$$n_z = n_z^{opt} = n \frac{p_z \sigma_{X|Z=z}}{\sum_{w=1}^{m} p_w \sigma_{X|Z=w}} \tag{25}$$

i.e. the so-called optimal sampling is performed. Therefore we have:

$$\begin{aligned} \sigma_X^2 &\ge \sum_{z=1}^{m} p_z \sigma_{X|Z=z}^2 = \sum_{z=1}^{m} \frac{n}{n_z^{prop}} p_z^2 \sigma_{X|Z=z}^2 \ge \\ &\ge \sum_{z=1}^{m} \frac{n}{n_z^{opt}} p_z^2 \sigma_{X|Z=z}^2 = \left(\sum_{z=1}^{m} p_z \sigma_{X|Z=z}\right)^2 \end{aligned} \tag{26}$$

For subsequent use let us define:

$$\sigma_{X,prop}^2 = \sum_{z=1}^{m} p_z \sigma_{X|Z=z}^2, \quad \sigma_{X,opt} = \sum_{z=1}^{m} p_z \sigma_{X|Z=z} \tag{27}$$

i.e. $\sigma_{X,prop}^2$ and $\sigma_{X,opt}^2$ are stratified variances of X for the proportional and optimal cases.

If n is fixed, then in order to find n_z^{prop}, $1 \le z \le m$, we only need to know Pr(Z=z), which is a frequent assumption regarding the stratified sampling technique. However, sometimes Pr(Z=z) are not known and have to be found, e.g. by simulation. On the other hand, n_z^{opt}, $1 \le z \le m$, can only be calculated if the conditional variances $\sigma^2_{X|Z=z}$ are known. As this is seldom the case, the usual practice consists in performing certain number of "pilot" simulations in order to find approximate values of $\sigma^2_{X|Z=z}$, necessary to compute n_z^{opt}, $1 \leq z \leq m$.

Stratified sampling appears to be well-fitted to the estimation problem considered in this paper, as both U and Q are dependent on Z defined as the number of the first component failed from the beginning of an operational cycle (i.e. from an instant τ_k, $k \geq 1$). Clearly, the value set of Z is equal to $\{1,...,m\}$ and

$$\Pr(Z = z) = \lambda_z/(\lambda_1 + \ldots + \lambda_m), \quad 1 \leq z \leq m \tag{28}$$

However, in our considerations U and Q are replaced with $U^* = U|Q\geq 1$ and $Q^* = Q|Q\geq 1$, as the condition "$Q\geq 1$" is necessary so that Lemma 1 can be applied. In consequence, Z has to be replaced with $Z^* = Z|Q\geq 1$, because stratification technique requires the use of the conditional expectations $E(U^*|Z^*=z)$ and $E(Q^*|Z^*=z)$, $1 \leq z \leq m$, and the conditioning random variable Z^* has to be defined on the same probability space as U^* and Q^*, i.e. $\{\omega: Q(\omega)\geq 1\}$. The evaluation of $E(U^*|Z^*=z)$ and $E(Q^*|Z^*=z)$ will be based on the following lemma:

Lemma 2

The conditional expectations of U^* and Q^* with respect to Z^* are given by:

$$E(U^* \mid Z^* = z) = E(U \mid Q \geq 1, Z = z), \quad E(Q^* \mid Z^* = z) = E(Q \mid Q \geq 1, Z = z) \tag{29}$$

Proof:

We will prove the first of the above equalities, the proof of the second one follows the same pattern. In view of the fact that U*=U and Z*=Z on {ω: Q(ω)≥1}, and Pr(·)/Pr(Q≥1) is the probability measure on {ω: Q(ω)≥1}, we have:

$$E(U^*|Z^*=z)=E(U^*,Z^*=z)/Pr(Z^*=z)=$$

$$=\frac{1}{Pr(Z^*=z)}\int_{\{Z^*=z,\,Q\geq 1\}} U^*(\omega)\frac{dP(\omega)}{Pr(Q\geq 1)} =$$

$$=\frac{1}{Pr(Z^*=z)}\int_{\{Z=z,\,Q\geq 1\}} U(\omega)\frac{dP(\omega)}{Pr(Q\geq 1)} = \tag{30}$$

$$=\frac{E(U,Z=z,Q\geq 1)}{Pr(Z^*=z)Pr(Q\geq 1)}=\frac{E(U|Z=z,Q\geq 1)Pr(Z=z,Q\geq 1)}{Pr(Z^*=z)Pr(Q\geq 1)}$$

and

$$Pr(Z^*=z)=Pr(Z=z|Q\geq 1)=\frac{Pr(Z=z,Q\geq 1)}{Pr(Q\geq 1)}=$$

$$=\frac{Pr(Q\geq 1|Z=z)Pr(Z=z)}{Pr(Q\geq 1)} \tag{31}$$

From (30) and (31) we obtain (29), which completes the proof.

The estimation of E(U*) and E(Q*) will be performed using their stratified estimators based on the following formulas derived directly from (29):

$$E(U^*) = \sum_{z=1}^{m} E(U \mid Q \geq 1, Z = z) \Pr(Z^* = z)$$

$$E(U^*) = \sum_{z=1}^{m} E(U \mid Q \geq 1, Z = z) \Pr(Z^* = z) \tag{32}$$

The probabilities Pr(Z*=z) (unlike Pr(Z=z)) are not given a priori, but can be computed from (31), while Pr(Q≥1) can be found from the following obvious equality:

$$\Pr(Q \geq 1) = \sum_{z=1}^{m} \Pr(Q \geq 1 \mid Z = z) \Pr(Z = z) \tag{33}$$

Thus, in order to evaluate E(U*) and E(Q*) we only need the approximate values of E(U|Q≥1, Z=z), E(Q|Q≥1, Z=z), and Pr(Q≥1|Z=z), which will be obtained by means of simulation.

We can now pass on to our main task, i.e. estimating the quotient E(U*) / E(Q*) by means of stratified sampling, where Z* is the stratification variable. In order to assess how many samples are necessary to accomplish this task we need the following "stratified" equivalent of Lemma 1:

Lemma 3

If X and Y are as in Lemma 1, $\hat{X}_{K,st}$ and $\hat{Y}_{L,st}$ are stratified estimators of μ_X and μ_Y, of sizes K and L respectively, and

$$\varepsilon_{\alpha,st} = \frac{2q_{1-\alpha/4}}{y_{min}^2} \max\left(\frac{\sigma_{X,st}\mu_Y}{\sqrt{K}}, \frac{\sigma_{Y,st}\mu_X}{\sqrt{L}}\right) \tag{34}$$

then, for sufficiently large K and L, $[\hat{X}_{K,st}/\hat{Y}_{L,st} - \varepsilon_{\alpha,st},\ \hat{X}_{K,st}/\hat{Y}_{L,st} + \varepsilon_{\alpha,st}]$ is a $1-\alpha$ confidence interval for μ_X/μ_Y.

The proof of Lemma 3 is conducted in the same way as that of Lemma 1, the only difference being that "stratified" standard deviations, i.e. $\sigma_{X,st}$ and $\sigma_{Y,st}$ are used in place of "raw" ones, i.e. σ_X and σ_Y. The immediate consequence of the above lemma is the following corollary: If ε is a given small number and

$$K \geq \left(\frac{2q_{1-\alpha/4}\sigma_{U^*,st}\mu_{Q^*}}{\varepsilon} \right)^2 \quad \wedge \quad L \geq \left(\frac{2q_{1-\alpha/4}\sigma_{Q^*,st}\mu_{U^*}}{\varepsilon} \right) \tag{35}$$

then taking at least K samples from U* and L samples from Q* allows to estimate μ_{U^*}/μ_{Q^*} with the accuracy at least ε (half-length of the confidence interval) at the given confidence level $1-\alpha$. In practice, μ_{U^*}, $\sigma_{U^*,st}$, μ_{Q^*}, and $\sigma_{Q^*,st}$ are replaced in (35) with the respective sample means or sample standard deviations.

We can now present (in outline) the algorithm for the estimation of E(U*)/E(Q*), applying the stratified sampling technique.

Algorithm 2

- For each z=1,...,m perform N_{pil} pilot simulations to find approximate values of $p_z^*=\Pr(Z^*=z)$, $\mu_{U^*|Z=z}$, $\sigma_{U^*|Z=z}^2$, $\mu_{Q^*|Z=z}$, and $\sigma_{Q^*|Z=z}^2$
- Compute minimal values of K and L fulfilling (35). E.g. for the proportional case:

$$K = (2q_{1-\alpha/4}\sigma_{U,prop}\mu_Q/\varepsilon)^2, \quad L = (2q_{1-\alpha/4}\sigma_{Q,prop}\mu_D/\varepsilon)^2 \tag{36}$$

- For each z=1,...,m compute K_z and L_z – the sizes of random samples from U*|Z*=z and Q*|Z*=z used to obtain the stratified estimators of μ_{U^*} and μ_{Q^*}. E.g. for the proportional case:

$$K_z = K\,p_z^*,\ L_z = L\,p_z^* \tag{37}$$

- Execute the main estimation procedure outlined below

Procedure 2 (the proportional case)

```
μU* = 0 ; μQ* = 0 ;
for z = 1,…,m do {
   j1 = 0 ; j2 = 0 ; μU*|Z=z = 0 ; μQ*|Z=z = 0 ;
   do {
      simulate one "truncated" operational cycle
      of X, i.e. begin from ez's failure, assuming that
      ez is the first failed component ;
      if (Q≥1) {
         if (j1 < Kz) {
            j1 = j1 + 1 ;
            μU*,z = μU*,z + (U – μU*|Z=z ) / j1 ;
         }
         if (j2 < Lz) {
            j2 = j2 + 1 ;
            μQ*,z = μQ*|Z=z + (Q – μQ*|Z=z ) / j2 ;
         }
      }
   } while (j1 < Kz  OR  j2 < Lz );
   μU* = μU* + pz* μU*|Z=z ;
   μQ* = μQ* + pz* μQ*|Z=z ;
}
return (μU* / μQ* );
```

As in Procedure 1, μ_{U*} and μ_{Q*} are updated based on (10) and (11).

In the optimal case $L_z = 0$ if $\sigma_{Q*,z} = 0$ which occurs if $Pr(Q=1|Z=z) = 1$ (e.g. for a consecutive system, $i = m$, $1 \le z \le m$). As the condition $j_2 < L_z$ is then not fulfilled in any cycle of the do-while loop, the command "if ($L_z = 0$) then $\mu_{Q*,z} = 1$" has to follow the do-while loop in order to customize Procedure 2 to the optimal case.

5. Estimation of other reliability parameters

Directly from the definition of $c^{(i)}$ it follows that

$$c^{(i)} = [a^{(i)} + b^{(i)}]^{-1} \tag{38}$$

provided that $b^{(i)}$ exists, i.e. $[B_1^{(i)} + \ldots + B_n^{(i)}]/n$ converges in probability to a constant value. The following lemma (proved in [Malinowski 2009]) defines the confidence limits for $c^{(i)}$.

Lemma 4

Let

$$\varphi_\alpha = \frac{\varepsilon_\alpha}{[a^{(i)} + r_{min}]^2} \tag{39}$$

where ε_α is given by (8). Then for sufficiently large K and L we have:

$$\Pr\left(\left|\frac{1}{a^{(i)}+\frac{\mu_{U*}}{\mu_{Q*}}}-\frac{1}{a^{(i)}+\frac{\hat{U}*_K}{\hat{Q}*_L}}\right|>\varphi_\alpha\right)\le\alpha \tag{40}$$

i.e. $[\,(a^{(i)}+\hat{U}*_K/\hat{Q}*_L)^{-1}-\varphi_\alpha,\ (a^{(i)}+\hat{U}*_K/\hat{Q}*_L)^{-1}+\varphi_\alpha\,]$ is a $1-\alpha$ confidence interval for $c^{(i)}$.

As in the case of Lemma 1, Lemma 4 can be converted into its "stratified" version by substituting σ_{U*} and σ_{Q*} with $\sigma_{U*,\,st}$ and $\sigma_{Q*,\,st}$ respectively. Based on the converted Lemma 4 we obtain the following corollary: If φ is a given small number and

$$K\ge\left[\frac{2q_{1-\alpha/4}\sigma_{U*,st}\mu_{Q*}}{\varphi(a^{(i)}+r_{min})^2}\right]^2 \wedge\ L\ge\left[\frac{2q_{1-\alpha/4}\sigma_{Q*,st}\mu_{U*}}{\varphi(a^{(i)}+r_{min})^2}\right]^2 \tag{41}$$

then $(a^{(i)}+\hat{U}*_{K,st}/\hat{Q}*_{L,st})^{-1}$ estimates $c^{(i)}$ with the accuracy at least φ (half-length of the confidence interval) at the given confidence level $1-\alpha$. Using (41), the algorithm estimating $c^{(i)}$, similar to Algorithm 2, can be constructed.

One more important reliability parameter characterizing repairable (renewable) systems is component or system availability. The most often used is the so called average availability defined as the percentage of component or system uptime over a long period of time. According to this definition we obtain the following formula

$$AV^{(i)}=\lim_{n\to\infty}\left(\frac{\sum_{j=1}^{n}A_j(i)}{\sum_{j=1}^{n}A_j(i)+B_j(i)}\right) \tag{42}$$

where $AV^{(i)}$ denotes the average availability of e_i. Clearly, if $b^{(i)}$ defined in (1) is a constant (see Lemma 2 in [Malinowski 2009]), then $AV^{(i)}$ is also a constant and

$$AV^{(i)} = \lim_{n\to\infty}\left(\frac{\sum_{j=1}^{n} A_j(i)}{n}\frac{n}{\sum_{j=1}^{n} A_j(i)+B_j(i)}\right) = \frac{a^{(i)}}{a^{(i)}+b^{(i)}} = \left(1+\frac{b^{(i)}}{a^{(i)}}\right)^{-1} \tag{43}$$

From (43) and Lemma 4 (converted to the stratified case) we obtain that if

$$\chi_\alpha = \frac{a^{(i)}\varepsilon_\alpha}{[a^{(i)}+r_{min}]^2} \tag{44}$$

with ε_α being given by (8), then for sufficiently large K and L $[(1+\hat{U}*_{K,st}/a^{(i)}\hat{Q}*_{L,st})^{-1}-\chi_\alpha,\ (1+\hat{U}*_{K,st}/a^{(i)}\hat{Q}*_{L,st})^{-1}+\chi_\alpha]$ is a $1-\alpha$ confidence interval for $AV^{(i)}$. We also have the following corollary: If χ is a given small number and

$$K \ge \left[2q_{1-\alpha/4}\sigma_{U*,st}\mu_{Q*}\frac{a^{(i)}}{\chi(a^{(i)}+r_{min})^2}\right]^2 \wedge\ L \ge \left[2q_{1-\alpha/4}\sigma_{Q*,st}\mu_{U*}\frac{a^{(i)}}{\chi(a^{(i)}+r_{min})^2}\right]^2 \tag{45}$$

then $(1+\hat{U}*_{K,st}/a^{(i)}\hat{Q}*_{L,st})^{-1}$ estimates $AV^{(i)}$ with the accuracy at least χ (half-length of the confidence interval) at the given confidence level $1-\alpha$. Using (45), the algorithm estimating $AV^{(i)}$, similar to Algorithm 2, can be constructed.

6. Exemplary numerical results

Several results obtained with Procedures 1 and 2 for the system in Fig. 1 are presented in Tables 1 and 2. It is assumed that $L_i^{(2)} = \infty$ with probability one for $1 \leq i \leq m$ (i.e. $G_i \equiv 0$), which means that components disconnected from e_0 cannot fail. Furthermore $L_i^{(1)}$ and R_i are exponentially distributed, mean time to failure is 500 h for a node, 250 h for a line. Mean time to repair is 10 h for a node, 5 h for a line. The time unit is one hour. K, K*, and T denote the number of all simulation cycles, the number of cycles with Q≥1, and the computing time respectively. V(·) and V_strat(·) are used to denote non-stratified and stratified variance, i.e. $\sigma_{\cdot}^2$ and $\sigma_{\cdot,st}^2$. The computations were carried out on a PC machine with an Intel Core 2 (2.14 GHz) processor.

Table 1

Examined component: 8 Accuracy: 0.02 Confidence level: 0.99	
Repair policy: FIFO, No. of repair teams: 1	Repair policy: LIFO, No. of repair teams: 1
Raw sampling E(U*)/E(Q*)=10.6077 E(Q*)=1.0178 V(U*)=152.3857 V(Q*)=0.0183 K=7,983,516, K*=1,994,304, T=1'25"	Raw sampling E(U*)/E(Q*)=9.9729 E(Q*)=1.0279 V(U*)=140.1585 V(Q*)=0.0327 K=7,494,940, K*=1,871,033, T=1'25"
Proportional sampling E(U*)/E(Q*)=10.6036 E(Q*)=1.0178 V_strat(U*)=119.2477 V_strat(Q*)=0.0184 K=6,287,391, K*=1,572,673, T=1'15"'	Proportional sampling E(U*)/E(Q*)=9.9596 E(Q*)=1.0293 V_strat(U*)=114.6449 V_strat(Q*)=0.0310 K=6,193,904, K*=1,544,858, T=1'15"
Optimal sampling E(U*)/E(Q*)=10.5976 E(Q*)=1.0174 V_strat(U*)=105.1993 V_strat(Q*)=0.0127 K=7,189,606, K*=1,384,276, T=1'20"	Optimal sampling E(U*)/E(Q*)=9.9670 E(Q*)=1.0292 V_strat(U*)100.7258 V_strat(Q*)=0.0218 K=7,106,720, K*=1,361,603, T=1'20"

Table 2

Examined component: 24 Accuracy: 0.02 Confidence level: 0.99	
Repair policy: FIFO, No. of repair teams: 1	Repair policy: FIFO, No. of repair teams: 2
Raw sampling E(U*)/E(Q*)=10.1233 E(Q*)=1.0394 V(U*)=171.9243 V(Q*)=0.0424 K=33,074,521, K*=14,665,709, T=6'10"	Raw sampling E(U*)/E(Q*)=7.0713 E(Q*)=1.0413 V(U*)=69.0699 V(Q*)=0.0508 K=13,619,352, K*=5,913,241, T=2'30"
Proportional sampling E(U*)/E(Q*)=10.1188 E(Q*)=1.0400 V_strat(U*)=141.3368 V_strat(Q*)=0.0412 K=27,483,419, K*=12,195,975, T=4'15"	Proportional sampling E(U*)/E(Q*)=7.0753 E(Q*)=1.0410 V_strat(U*)=62.0311 V_strat(Q*)=0.0485 K=12,475,669, K*=5,414,678, T=2"
Optimal sampling E(U*)/E(Q*)=10.1263 E(Q*)=1.0393 V_strat(U*)=130.9074 V_strat(Q*)=0.0373 K=30,516,923, K*=11,186,840, T=4'40"	Optimal sampling E(U*)/E(Q*)=7.0803 E(Q*)=1.0408 V_strat(U*)=60.2416 V_strat(Q*)=0.0332 K=12,895,815, K*=5,193,306, T=2"

As follows from the presented results, stratified sampling, applied for the considered system, does not significantly reduce the number of computations. The reason lies in relatively small differences between $\sigma_{U^*,z}$ and $\sigma_{Q^*,z}$, $1 \le z \le m$, not shown in the tables. Generally, stratified sampling gives good results when conditional variances, as defined by (19), differ considerably.

It can be observed that the computing time strongly depends on the number of repair teams, the reason being that operational cycles are longer for smaller number of repair teams. A component's distance from e_0 also impacts the computing time which is shorter for less distant components. On the other hand, the computing effort does not practically depend on the repair policy applied.

As could be expected, $b^{(i)}$ is more sensitive to the type of repair policy in case of components less distant from e_0. Clearly, if components located close to the source are repaired before those located further away (LIFO policy), then the former ones are likely to become reconnected to e_0 more quickly than in case of FIFO policy.

It is interesting that, paradoxically, total number of simulations in the optimal case can be greater than the respective number in the proportional case (e.g. compare the results in the first column of Table 2). For explanation note that the numbers K_z or L_z, $1 \leq z \leq m$, as given by suitably modified (23) or (25), differ to a greater extent in the former than in the latter case. Also note that for each z the total number of simulation cycles in Procedure 2 is proportional to $\max(K_z, L_z)$ – the number of cycles with $Q \geq 1$. Therefore the bigger differences among K_z or L_z lead to the larger number of all cycles.

6. Bibliography

R. Billinton, R. N. Allan: *Reliability Evaluation of Engineering Systems*, Springer 1992

R. Billinton, R. N. Allan: *Reliability Evaluation of Power Systems*, Plenum Press 1996

R. Billinton, W. Li: *Reliability Assessment of Electrical Power Systems Using Monte Carlo Methods*, Springer 1994

A.A. Borovkov: *Mathematical Statistics*, Gordon and Breach Science Publishers 1998

R. E. Brown: *Electric Power Distribution Reliability*, CRC Press 2002

W. Feller: *An Introduction to Probability Theory and Its Applications*, Wiley 1968

P. Glasserman: *Monte Carlo Methods in Financial Engineering*, Springer 2003

J. Malinowski: *Simulating failure-repair process and evaluating reliability parameters for single-source multiple-sink commodity transportation network with stochastically dependent components*, IBS PAN 2009

J. Malinowski: *Accuracy analysis for reliability parameters estimation of a network system with non-independent components*, Reliability, Risk and Safety – Back to the Future, pp. 990-997, CRC Press 2010

J.E. Ramirez-Marquez, G. Levitin: *Algorithm for estimating reliability confidence bounds of multi-state systems*, Reliability Engineering and System Safety, Vol. 93, pp. 1231-1243, 2008

S.E Ridgen, A.P. Basu: *Statistical Methods for the Reliability of Repairable Systems*, Wiley 2000