INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Monte Carlo Simulation of the Flow Network Reliability using Importance and Stratified Sampling

Stéphane Bulteau, Mohamed El Khadiri

N° 3122

Mars 1997

THÈME 1



ISSN 0249-6399



A Monte Carlo Simulation of the Flow Network Reliability using Importance and Stratified Sampling

Stéphane Bulteau*, Mohamed El Khadiri †

Thème 1 — Réseaux et systèmes
Projet Model

Rapport de recherche n ° 3122 — Mars 1997 — 17 pages

Abstract: We consider the evaluation of the flow network reliability parameter. Because the exact evaluation of this parameter has exponential time complexity, simulation methods are used to derive an estimate. In this paper, we use the state space decomposition methodology of Doulliez and Jamoulle for constructing a new simulation method which combines the importance and the stratified Monte Carlo principles. We show that the related estimator belongs to the variance-reduction family. By numerical comparisons, we illustrate the interest of our method when compared to the previous simulation methods based on the same decomposition.

Key-words: Flow network, maximum st-flow, Monte Carlo simulation, variance-reduction, importance and stratified sampling

(Résumé : tsvp)

* Email: sbulteau@irisa.fr
† Email:khadiri@irisa.fr

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : (33) 02 99 84 71 00 – Télécopie : (33) 02 99 84 71 71

Une méthode de Monte Carlo pour le problème de la fiabilité d'un réseau de transport utilisant un échantillonnage préférentiel et stratifié

Résumé : Nous considérons le problème de l'évaluation de la fiabilité d'un réseau de transport stochastique. Le fait que l'évaluation exacte de ce paramètre soit de complexité temporelle exponentielle, conduit à utiliser des simulations de Monte Carlo pour l'estimer. Dans ce papier, nous exploitons la procédure de décomposition de l'espace d'états de Doulliez et Jamoulle pour construire une nouvelle méthode de simulation, basée sur l'échantillonnage stratifié et l'échantillonnage préférentiel. Nous démontrons l'appartenance de cette méthode à la famille des méthodes de réduction de la variance et nous la comparons à une précédente méthode basée sur la même procédure de décomposition. Nous illustrons par un exemple l'intérêt de notre méthode par rapport à une précédente méthode basée sur la même décomposition.

Mots-clé : Réseau de transport, st-flot maximal, simulation de Monte Carlo, reduction de la variance, échantillonnage préférentiel et stratifié.

1 Introduction

A basic mission of a flow network is the establishment of a flow between a source node s and a sink node t that meets or exceeds a fixed demand d. In the stochastic case, the network components have random capacities and the success of this mission is a random event. Systems that can be regarded as flow networks are electric-power transmission, distribution systems, transportation networks and computer networks. The probability of successful mission is a performance measure, often called the *flow network reliability measure*.

Several papers have been devoted to its evaluation when nodes do not limit flow transmission and when arcs capacities are discrete, multi-valued and statistically independent random variables [7, 10, 14, 15]. In this case, one of the methods to compute exactly the parameter under consideration is based on the state space decomposition methodology of Doulliez and Jamoulle [7]. It starts by decomposing the state space into non-overlapping operating set, failed set, and undetermined sets. Each undetermined set is used as input of subsequent similar decomposition. The recursive process terminates when all generated sets are classed operating or failed. The probability that the random vector state belongs to any generated operating set is easy to compute and the flow network reliability parameter is the sum of these probabilities.

When all arcs have only two possible capacities 1 or 0 and the demand is $d = 1$, the problem becomes the source-terminal reliability problem [5], [6], [9] which is NP-hard [2]. Consequently, the general case considered here is also an NP-hard problem. This implies that the computational time will be prohibitive when the network size is large [8]. Then Monte Carlo approaches are alternatives allowing, in a reasonable time, the evaluation of large networks.

For a fixed sample size K, the estimation of the flow network reliability parameter by the standard estimator is the frequency of operating capacity vectors in the set of K independent trials drawn from the whole state space of the random capacity vector. In order to obtain more efficient estimator with the same sample size K, in [12], Fishman and Shaw stop the exact algorithm (using recursively the decomposition procedure of Doulliez and Jamoulle) after a pre-fixed number of calls to the decomposition procedure and exploit some resulting informations in the simulation context: the generated operating and failed sets are used to deduce a lower bound and an upper bound on the reliability parameter, the sampling is reduced to the union of the resulting undetermined subsets and the K trials are distributed on these sets. Hence, each undetermined set constitutes a stratum and the method is called a stratified sampling technique.

In this technique, the trials within each stratum is accomplished by the standard sampling technique. The aim of our work is to show that if we use in each stratum the importance sampling method proposed in [4], we obtain more accurate estimator. The used importance sampling technique allows to transform the sampling in each stratum into the sampling in a smaller set, and this recursively, until it is not possible to accomplish new decompositions. We then avoid the use of the standard sampling technique during the simulation process. The paper is organized as follows. The following section introduces some

notations and the model definition. In section 3, we recall the standard estimator. In section 4, we resume the exact algorithm of Doulliez and Jamoulle and we recall the stratified estimator proposed by Fishman and Shaw [12]. We also present our stratified Monte Carlo technique. Section 5 is devoted to numerical comparisons and Section 6 to some conclusions.

2 Notations and model definition

We resume in this section the model definition and some notations. For ease of explanation, additional definitions and notation will be given in adequate sections.

We will denote $G = (V, A, \vec{\mathcal{C}}, s, t, d)$, the flow network where, V is the set of nodes, $A = \{e_1, \ldots, e_a\}$, the set of arcs, s is the source node, t, the sink node and d, the demand required at node t.

- For each arc $e_j \in A$,
  - $\mathcal{C}_j$ denotes the random discrete capacity of arc e_j,
  - $0 \leq c_{j1} < c_{j2} < \ldots < c_{jn_j} < +\infty$ are the n_j possible values of the random variable $\mathcal{C}_j$,
  - and $\Omega_j = \{c_{j1}, \ldots, c_{jn_j}\}$ is the state space of the random variable $\mathcal{C}_j$,
  - p_{jn} is the probability that $\mathcal{C}_j$ has capacity c_{jn} in Ω_j.
- $\vec{\mathcal{C}} = (\mathcal{C}_1, \ldots, \mathcal{C}_a)$ denotes the random state vector of the network G. A value $\vec{C} = (c_1, \ldots, c_a)$ of this vector is called a capacity vector.
- $\Omega = \bigotimes_{j=1}^{a} \Omega_j$ represents the network state space (the state space of random variable $\vec{\mathcal{C}}$).
- $v(c_1, \ldots, c_a)$ is the value of maximum st-flow when the arc e_j has capacity c_j, $e_j \in A$.
- We denote Φ the non-decreasing structure function of the network G defined for all capacity vectors $\vec{C}$ by
$$\Phi(\vec{C}) = \begin{cases} 1 & \text{if } v(\vec{C}) \geq d \\ 0 & \text{otherwise.} \end{cases}$$
- A set $R \subseteq \Omega$ is *rectangular* if and only if there is a lower and an upper vector of capacities $\alpha(R) = (\alpha_1(R), \ldots, \alpha_a(R))$ and $\beta(R) = (\beta_1(R), \ldots, \beta_a(R))$, respectively, in Ω such that every capacity vector $\vec{C} = (c_1, \ldots, c_a)$ with $\alpha_j(R) \leq c_j \leq \beta_j(R)$ for all arcs e_j, belongs to R and R contains only those vectors [7].
- For any rectangular set $R \subseteq \Omega$, we associate the following notations and definitions:

- $\vec{\mathcal{C}}\ (R)$: the random vector having the same distribution as the conditional distribution of $\vec{\mathcal{C}}$ given that $\vec{\mathcal{C}} \in R$,
- $G(R) = (V, A, \vec{\mathcal{C}}\ (R), s, t, d)$: the network that differs from G only by its random state vector which is equal to $\vec{\mathcal{C}}\ (R)$,
- $\mathcal{B}(R) = \Phi(\vec{\mathcal{C}}\ (R))$,
- $g(R) = \mathrm{E}\{\mathcal{B}(R)\} = \mathrm{E}\left\{\Phi(\vec{\mathcal{C}}\ (R))\right\} = \Pr\left\{\Phi(\vec{\mathcal{C}}\ (R)) = 1\right\} = \mathrm{E}\left\{\Phi(\vec{\mathcal{C}}) \mid \vec{\mathcal{C}} \in R\right\}$: the flow network reliability of the network $G(R)$.

- $g(\Omega) = \mathrm{E}\left\{\Phi(\vec{\mathcal{C}})\right\} = \Pr\left\{\Phi(\vec{\mathcal{C}}) = 1\right\}$ is the considered reliability parameter of the flow network G under consideration.

3 The standard Monte Carlo method

For a sample size K and a rectangular set $R \subseteq \Omega$, the estimation of the reliability parameter $g(R)$, of the network $G(R)$, by the unbiased standard Monte Carlo estimator is the frequency of operating vectors in the set of K independent trials drawn from R using the distribution function of $\vec{\mathcal{C}}\ (R)$. More formally, this estimator is the sample mean,

$$\widehat{\Phi}(K, R) = \frac{1}{K} \sum_{i=1}^{K} \Phi(\vec{\mathcal{C}}^{(i)}(R)), \tag{1}$$

based on K independent trials $\Phi(\vec{\mathcal{C}}^{(1)}(R)), \ldots, \Phi(\vec{\mathcal{C}}^{(K)}(R))$ of the random variable $\mathcal{B}(R) = \Phi(\vec{\mathcal{C}}\ (R))$.

By the independence of the K trials and since $\Phi(\vec{\mathcal{C}}\ (R))$ is a Bernouilli random variable, the variance of the above estimator is

$$\mathrm{Var}\left\{\widehat{\Phi}(\vec{\mathcal{C}}\ (R))\right\} = \frac{\mathrm{Var}\left\{\Phi(\vec{\mathcal{C}}\ (R))\right\}}{K} = \frac{g(R)(1 - g(R))}{K}. \tag{2}$$

For a given sample size K, an estimate of $g(\Omega)$ by a standard estimator can be obtained from the following function with the parameter R equal to Ω. This function returns, for any rectangular subset R of Ω, an estimate of $g(R)$ by a sample mean based on K independent trials of $\Phi(\vec{\mathcal{C}}\ (R))$.

Function *StandardSampling*(R, K)

1. Initialization: $Est := 0$.
2. For each experiment $k := 1, \ldots, K$ do
   2.1. For each link $j \in A$ accomplish a trial c_j of $\mathcal{C}_j(R)$.
   2.2. Compute $v(c_1, c_2, \ldots, c_a)$.
   2.3. If $v(c_1, c_2, \ldots, c_a) \geq d$ then $Est := Est + 1$.
3. Return(Est/K)

The Ford-Fulkerson flow-augmenting method can be used to compute $v(c_1, c_2, \ldots, c_a)$, the maximum st-flow corresponding to the capacity vector $(c_1, c_2, \ldots, c_a)$. A method to generate a trial of a discrete random variable can be found for instance in [11].

The well known drawback of the standard Monte Carlo sampling is the large sample size K required to obtain sufficiently small variance and reasonable relative error [9]. In the following section, we will interest to the construction of Monte Carlo estimators based on the state space decomposition procedure of Doulliez and Jamoulle. These estimators have smaller variances than the standard estimator, for a fixed sample size K. This implies that they belong to the variance-reduction family, leading to more accurate estimates than the standard one [13].

4 A Monte Carlo methods based on the decomposition procedure of Doulliez and Jamoulle

Before we present Monte Carlo methods based on the decomposition of Doulliez and Jamoulle, we recall briefly its exploitation for computing exactly the reliability parameter $g(R)$, for any rectangular subset R of Ω. We also give a lemma that we consider as a start point for constructing stratified sampling methods treated here.

4.1 Basics

To evaluate $g(R)$, the exact algorithm based on the decomposition of Doulliez and Jamoulle [1, 7], starts by decomposing the set R into non-overlapping operating subset, failed subset and undetermined rectangular subsets. Each undetermined set is used as input of subsequent similar decomposition. This process terminates when all generated sets are classed operating or failed. The probability that the random vector state belongs to any generated operating set is easy to compute and the flow network reliability parameter is the sum of these probabilities [1, 7]. More formally, we have

$$g(R) = \sum_{i=1}^{T(R)} \Pr\left\{\vec{\mathcal{C}} \in W_i(R)\right\} \tag{3}$$

where $T(R)$ is the number of generated operating rectangular subsets and $W_i(R)$, $1 \leq i \leq T(R)$, are these sets.

In order to compute $g(\Omega)$, it suffices to replace R by Ω in formula (3). Because the exact evaluation is an NP-Hard problem and the decomposition procedure has polynomial time complexity [12], the exponential complexity of the exact algorithm based on (3) results from the exponential growth of the number of calls to the decomposition procedure, or equivalently the number of generated operating rectangular subsets, as a function of the network size. However, if we are limited to M decompositions, due to the execution time, we can deduce a lower and an upper bounds on the considered reliability parameter and exploit them in the simulation context [4, 12].

The following lemma expresses the reliability $g(R)$ ($R \subseteq \Omega$) as a function of bounds on this parameter and the exact reliabilities associated to undetermined subsets of R resulting from a number M of decompositions, when M is smaller than the total number of decompositions needed to compute exactly $g(\Omega)$. This will be useful in the construction of stratified simulation methods in the next subsections.

For the rest of this paper, we will need the following definitions and notation (valid for any rectangular set $R \subseteq \Omega$ and all methods):

- $T(R)$ is the number of calls to the decomposition procedure to compute $g(R)$ by formula (3)
- For a fixed integer $M \leq T(\Omega)$ and for any rectangular set $R \subseteq \Omega$,
  - $W(R)$: the subset of vectors classed operating in the M first decompositions for computing $g(R)$ by formula (3)
  - $g_l(R) = \Pr\left\{\vec{\mathcal{C}}\ (R) \in W(R)\right\}$: a lower bound on $g(R)$
  - $F(R)$: the subset of vectors classified failed in these decompositions
  - $h(R)$: the number of undetermined rectangular subsets not yet decomposed
  - $S(R) = \{U_1(R), \ldots, U_{h(R)}(R)\}$: the set of these undetermined subsets
  - $U(R) = \bigcup_{h=1}^{h(R)} U_h(R)$
  - $\pi_h(R) = \Pr\left\{\vec{\mathcal{C}}\ (R) \in U_h(R)\right\}$, for $h \in \{1, \ldots, h(R)\}$
  - $g_u(R) = 1 - \Pr\left\{\vec{\mathcal{C}}\ (R) \in F(R)\right\} = 1 - \sum_{h=1}^{h(R)} \pi_h(R)$: an upper bound on $g(R)$
  - If $S(R) \neq \emptyset$ ($h(R) \neq 0$), $\widetilde{U}(R)$ denotes the random variable defined in the set $S(R)$ by

$$\begin{aligned} \Pr\left\{\widetilde{U}(R) = U_h(R)\right\} &= \Pr\left\{\vec{\mathcal{C}}\ (R) \in U_h(R) \mid \vec{\mathcal{C}}\ (R) \in U(R)\right\} \\ &= \pi_h(R)/(g_u(R) - g_l(R)), \qquad \text{for } h \in \{1, \ldots, h(R)\}. \end{aligned}$$

Lemma 4.1 *For the fixed integer $M \leq T(\Omega)$ and for any rectangular set $R \subseteq \Omega$, we have the following transformation*

$$g(R) = g_l(R) + \sum_{h=1}^{h=h(R)} \pi_h(R) g(U_h(R)). \tag{4}$$

Proof.

As the sets $W(R)$, $F(R)$ and the undetermined sets $U_h(R)$, $h \in \{1, \ldots, h(R)\}$, form a partition of R, the total expectation theorem gives

$$\begin{aligned} g(R) = \mathrm{E}\{\mathcal{B}(R)\} &= \mathrm{E}\left\{\mathcal{B}(R) \mid (\vec{\mathcal{C}}\ (R) \in W(R))\right\} \Pr\left\{\vec{\mathcal{C}}\ (R) \in W(R)\right\} \\ &+ \mathrm{E}\left\{\mathcal{B}(R) \mid (\vec{\mathcal{C}}\ (R) \in F(R))\right\} \Pr\left\{\vec{\mathcal{C}}\ (R) \in F(R)\right\} \\ &+ \sum_{h=1}^{h=h(R)} \mathrm{E}\left\{\mathcal{B}(R) \mid (\vec{\mathcal{C}}\ (R) \in U_h(R))\right\} \Pr\left\{\vec{\mathcal{C}}\ (R) \in U_h(R)\right\}. \end{aligned}$$

Since $W(R)$ is an operating set and $F(R)$ is a failed set, we have

$$\mathrm{E}\left\{B(R) \mid (\vec{\mathcal{C}}\ (R) \in W(R))\right\} = 1$$

and

$$\mathrm{E}\left\{B(R) \mid (\vec{\mathcal{C}}\ (R) \in F(R))\right\} = 0.$$

We then obtain

$$g(R) = \mathrm{E}\{B(R)\} = g_l(R) + \sum_{h=1}^{h=h(R)} g(U_h(R)) \pi_h(R). \tag{5}$$

■

4.2 The previous stratified Monte Carlo method based on the decomposition procedure of Doulliez & Jamoulle

If we replace R by Ω in equality (5), we obtain

$$g(\Omega) = g_l(\Omega) + \sum_{h=1}^{h=h(\Omega)} \pi_h(\Omega) g(U_h(\Omega)). \tag{6}$$

Instead of drawing the K trials from Ω using the distribution function of $\vec{\mathcal{C}}=\vec{\mathcal{C}}$ (Ω), as in the standard procedure presented in section 3, equality (6) leads the authors in [12] to draw the K samples from the subset

$$U(\Omega) = \bigcup_{h=1}^{h=h(\Omega)} U_h(\Omega).$$

Moreover, they propose to distribute the K trials on the undetermined sets. Hence, each of them constitutes a stratum and the method is called a stratified sampling one. The number K_h of trials drawn from each $U_h(\Omega)$ is fixed proportional to the probability that $\vec{\mathcal{C}}$ belongs to $U_h(\Omega)$ given that $\vec{\mathcal{C}}$ belongs to $U(\Omega)$. The reader can see [12] for a discussion about the rules to distribute the K trials.

More formally, they suggest to use the following unbiased estimator

$$\widetilde{\mathcal{F}} = g_l(\Omega) + \sum_{h=1}^{h=h(\Omega)} \pi_h(\Omega)\widehat{\Phi}(K_h, U_h(\Omega)) \tag{7}$$

where

$$\begin{aligned} K_h &= K\Pr\left\{\widetilde{U}(\Omega) = U_h(\Omega)\right\} \\ &= K\pi_h(\Omega)/(g_U(\Omega) - g_l(\Omega)), \qquad \text{for } h = 1, \ldots, h(\Omega). \end{aligned}$$

and $\widehat{\Phi}(K_h, U_h(\Omega))$ is the standard unbiased estimator of $g(U_h(\Omega))$ based on K_h independent trials of $\vec{\mathcal{C}}$ $(U_h(\Omega))$ (see definition (1)).

It is showed in [12] that the variance of this estimator is

$$\mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\} = (g(\Omega) - g_l(\Omega))(g_u(\Omega) - g(\Omega))/K - \sum_{h=1}^{h=h(\Omega)} \pi_h(\Omega)(g(U_h(\Omega)) - g(U(\Omega))^2/K. \tag{8}$$

Then

$$\begin{aligned} \mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\} &\leq (g(\Omega) - g_l(\Omega))(g_u(\Omega) - g(\Omega))/K \\ &\leq g(\Omega)(1 - g(\Omega))/K. \end{aligned}$$

By equality (2), the value $g(\Omega)(1 - g(\Omega))/K$ corresponds to the variance of the standard estimator of $g(\Omega)$ and

$$\mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\} \leq \mathrm{Var}\left\{\widehat{\Phi}(\vec{\mathcal{C}}\ (\Omega))\right\}.$$

The last inequality implies that $\widetilde{\mathcal{F}}$ offers more accurate estimates than the standard one.

For a given sample size K, an estimate of $g(\Omega)$ by this sampling strategy can be obtained from the following function with the parameter R equal to Ω. In step 0 of this function we use the procedure Bounds$(R, M, g_l(R), g_u(R), h(R), S(R), \{\pi_1(R), \ldots, \pi_{h(R)}(R)\})$ that computes, for a given rectangular subset R of Ω and a fixed integer M of decompositions, the bounds $g_l(R)$, $g_u(R)$, the number $h(R)$ of undetermined rectangular subsets, the set of undetermined sets $S(R) = \{U_1(R), \ldots, U_{h(R)}(R)\}$, the set of associated weights $\{\pi_1(R), \ldots, \pi_{h(R)}(R)\}$. If $T(R) \leq M$, this procedure gives $g_l(R) = g_u(R) = g(R)$ and $h(R) = 0$. The reader can see [7, 12], for details about the computing process of the procedure Bounds(), even if this is not necessary to understand our work.

Function $StratifiedSampling1(K, R)$

0. Preliminary step: $Bounds(R, M, g_l(R), g_u(R), h(R), S(R), \{\pi_1(R), \ldots, \pi_{h(R)}(R)\}))$
1. Initialization: $Est := 0$
2. For $h := 1$ to $h(R)$ do
   2.1. Evaluate the size K_h: $K_h := K \times \pi_h(R)/(g_u(R) - g_l(R))$
   2.2. Compute the standard estimation of $g(U_h(R))$: $Est_h := StandardSampling(U_h(R), K_h)$
   2.3. Add the contribution of the stratum h to Est:
   $Est := Est + \pi_h(R) \times Est_h$
3. Return the estimate of $g(R)$: $return(g_l(R) + Est)$

4.3 A new stratified sampling estimator

In the above algorithm, the estimation of each reliability parameter $g(U_h(\Omega))$ associated to the stratum $U_h(\Omega)$, is accomplished by the standard sampling. If these reliabilities are estimated by estimators that belong to the variance-reduction family, we will obtain more accurate estimator than $\widetilde{\mathcal{F}}$ (7). As a consequence, we propose to replace the standard estimator by the recursive importance sampling one presented in [3, 4]. For any rectangular subset R of Ω, this estimator is a sample mean based on the random variable $\mathcal{Z}(R)$ given in the next proposition. When we compute a trial of $\mathcal{Z}(R)$, we transform the sampling in R into the sampling in one undetermined subset among the $h(R)$ undetermined subsets resulting from M decompositions of R. The selection of this undetermined subset results from a trial of the random variable $\widetilde{U}(R)$ defined in subsection 4.1. We then use recursively this transformation until it is called on an undetermined subset such that its reliability can be computed exactly by a number of decompositions smaller than or equal to a fixed integer M. An adequate combination of this reliability and other probabilities collected during the recursive sampling process constitutes a trial of $\mathcal{Z}(R)$. By this technique, we avoid the use of the standard sampling technique during the simulation process.

4.3.1 Theoretical developments

In the following proposition, we recall the random variable $Z(R)$, used to construct the recursive importance sampling estimator, and we define the new stratified estimator.

Proposition 4.2 *For a fixed integer $M \leq T(\Omega)$ and for any rectangular set $R \subset \Omega$, let us define the recursive random variable $\mathcal{Z}(R)$ by:*

$$\mathcal{Z}(R) = \begin{cases} g(R) & \text{if } T(R) \leq M \\ g_l(R) + (g_u(R) - g_l(R))\mathcal{Z}(\widetilde{U}(R)) & \text{otherwise,} \end{cases} \tag{9}$$

where all parameters are defined as in lemma (4.1).
Let

$$\widehat{\mathcal{Z}}(K_h, U_h(\Omega)) = \frac{1}{K_h} \sum_{i=1}^{K_h} \mathcal{Z}^{(i)}(U_h(\Omega))$$

the sample mean based on K_h independent trials using the distribution function of $\mathcal{Z}(U_h(\Omega))$. Then, the estimator

$$\widetilde{\mathcal{Z}} = g_l(\Omega) + \sum_{h=1}^{h=h(\Omega)} \widehat{\mathcal{Z}}(K_h, U_h(\Omega))\pi_h(\Omega) \tag{10}$$

verifies

$$\mathrm{E}\left\{\widetilde{\mathcal{Z}}\right\} = g(\Omega) = \mathrm{E}\left\{\widetilde{\mathcal{F}}\right\} \tag{11}$$

and

$$\mathrm{Var}\left\{\widetilde{\mathcal{Z}}\right\} \leq \mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\} \leq \mathrm{Var}\left\{\widehat{\Phi}(\vec{\mathcal{C}})\right\} \tag{12}$$

Proof. We can prove with a recursion on $T(R)$ that, for any rectangular set R,

$$\mathrm{E}\{\mathcal{Z}(R)\} = \mathrm{E}\{\mathcal{B}(R)\} = g(R) \tag{13}$$

and

$$\mathrm{Var}\{\mathcal{Z}(R)\} \leq (g(R) - g_l(R))(g_u(R) - g(R)) \leq g(R)(1 - g(R)) = \mathrm{Var}\{\mathcal{B}(R)\} \tag{14}$$

Equality (13) implies that a sample mean based on $\mathcal{Z}(R)$ is an unbiased estimator of $g(R)$ and equality (14) implies that this estimator leads to more accurate estimates than the standard one. The reader can see [4] for the detailed proof of equalities (13) and (14).
By replacing R by $U_h(\Omega)$ in equality (13), we obtain that $\mathrm{E}\left\{\widehat{\mathcal{Z}}(K_h, U_h(\Omega))\right\} = g(U_h(\Omega))$ and then,

$$\begin{aligned} \mathrm{E}\left\{\widetilde{\mathcal{Z}}\right\} &= g_l(\Omega) + \sum_{h=1}^{h=h(\Omega)} \pi_h(\Omega)\mathrm{E}\left\{\widehat{\mathcal{Z}}(K_h, U_h(\Omega))\right\} \\ &= g_l(\Omega) + \sum_{h=1}^{h=h(\Omega)} \pi_h(\Omega) g(U_h(\Omega)) = g(\Omega) \end{aligned}$$

from formula (6).
For the variance, by applying the result (14) with $U_h(\Omega)$, we obtain

$$\mathrm{Var}\left\{\widehat{\mathcal{Z}}(K_h, U_h(\Omega))\right\} = \frac{\mathrm{Var}\left\{\mathcal{Z}(U_h(\Omega))\right\}}{K_h} \leq \frac{\mathrm{Var}\left\{\mathcal{B}(U_h(\Omega))\right\}}{K_h} = \mathrm{Var}\left\{\widehat{\Phi}(K_h, U_h(\Omega))\right\}.$$

As

$$\mathrm{Var}\left\{\widetilde{\mathcal{Z}}\right\} = \sum_{h=1}^{h=h(\Omega)} \pi_h^2(\Omega)\mathrm{Var}\left\{\widehat{\mathcal{Z}}(K_h, U_h(\Omega))\right\}$$

and

$$\mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\} = \sum_{h=1}^{h=h(\Omega)} \pi_h^2(\Omega)\mathrm{Var}\left\{\widehat{\Phi}(K_h, U_h(\Omega))\right\}$$

We deduce that

$$\mathrm{Var}\left\{\widetilde{\mathcal{Z}}\right\} \leq \mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\}.$$

■

Result (11) implies that $\widetilde{\mathcal{Z}}$ is an unbiased estimator of $g(\Omega)$. Result (12) implies that the use of the recursive importance sampling to estimate the contributions of the strata leads to more accurate estimator than the previous stratified estimator proposed in [12].

4.3.2 Algorithmic description of the new simulation algorithm

The function *ImportanceSampling*() that gives a trial of the random variable $\mathcal{Z}(R)$ defined by the recursive formula (9) can be described as follows:

Function $ImportanceSampling(R)$

1. Check end recursion condition:
   1.1. $Bounds(R, M, g_l(R), g_u(R), h(R), S(R), \{\pi_1(R), \ldots, \pi_{h(R)}(R)\}))$
   1.2. If $(h(R) = 0)$ return$(g_l(R))$
2. Accomplish a trial $R' \in S(R)$ by using the distribution function of $\widetilde{U}(R)$
3. Recursive call: return$(g_l(R) + (g_u(R) - g_l(R)) \times ImportanceSampling(R'))$

For a given sample size K, an estimate of $g(\Omega)$ by the proposed strategy can be obtained from the function
$StratifiedSampling2()$ with the parameters K and Ω. As in the function $StratifiedSampling1()$, the new function uses the procedure $Bounds()$.

Function $StratifiedSampling2(K, R)$

0. Preliminary step: $Bounds(R, M, g_l(R), g_u(R), h(R), S(R), \{\pi_1(R), \ldots, \pi_{h(R)}(R)\})$
1. Initialization: $Est := 0$
2. For $h := 1$ to $h(R)$ do
   2.1. Evaluate the size K_h: $K_h := K \times \pi_h(R)/(g_u(R) - g_l(R))$
   2.2. Compute the estimation of $g(U_h(R))$ by a sample mean based on K_h trials of $\mathcal{Z}(U_h(R))$:
      2.2.1. $Est_h := 0$
      2.2.2. For $k := 1$ to K_h do $Est_h := Est_h + ImportanceSampling(U_h(R))$
      2.2.3. $Est_h := Est_h/K_h$
   2.3. Add the contribution of the stratum h to Est: $Est := Est + \pi_h(R) \times Est_h$
3. Return the estimate of $g(R)$: $return(g_l(R) + Est)$

5 Numerical illustrations

We illustrate our methodology by evaluating the reliability of the network with 10 nodes and 25 arcs represented in Figure 1 [12]. For each arc j, the random discrete capacity $\mathcal{C}_j$ is a Bernoulli random variable that has value 0 with probability q and value c_j (on the arc j of Figure 1) with probability $p = 1 - q$. We consider five different demands $d = 20$, 36, 50, 60 and 71, and the probability $p = 0.9$. When the capacity of every arc j is fixed at its upper value c_j, the maximum st-flow is equal to 71. Consequently a demand greater than 71 at the sink node t can not be satisfied. Then, such demands are not to be considered. We only consider the case, M=1. So, the procedure Bounds() is called with M=1.

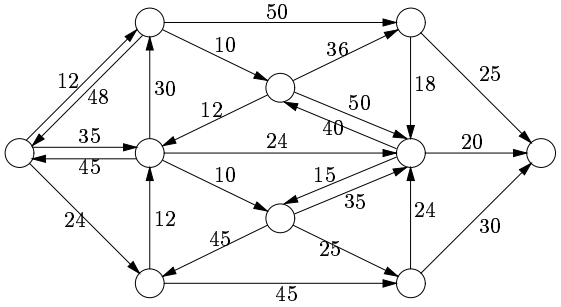


Figure 1: The network used for numerical results in Tables 1 and 2

In Table 1, we present at Columns 2 and 3 the exact values of $g(\Omega)$ and their estimates by $\widetilde{\mathcal{Z}}$ with $K = 2^{16}$. It can be observed that each estimate is close to the corresponding $g(\Omega)$. The exact algorithm of Doulliez and Jamoulle has been used to compute $g(\Omega)$ and $g_h(\Omega)$, for $h = 1, \ldots, h(\Omega)$. These values serve to tabulate, at Column 4, the variance of $B(\Omega)$, which is equal to $g(\Omega)(1 - g(\Omega))$ and, at Column 5, the variance of $\widetilde{\mathcal{F}}$, evaluated by equation (8). Since the variance of $\widetilde{\mathcal{Z}}$ is unknown, its unbiased estimator [13]

$$\widehat{V}_{\widetilde{\mathcal{Z}}} = \sum_{h=1}^{h=h(\Omega)} \pi_h^2(\Omega) \frac{1}{K_h - 1} \sum_{i=1}^{K_h} (\widehat{\mathcal{Z}}_h - {\mathcal{Z}_h}^{(i)})^2 \tag{15}$$

is used to obtain estimates of $\mathrm{Var}\left\{\widetilde{\mathcal{Z}}\right\}$ at Column 6.

d	$g(\Omega)$	$\widetilde{\mathcal{Z}}$	$\mathrm{Var}\left\{\widehat{\Phi}(\vec{\mathcal{C}})\right\}$	$\mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\}$	$\mathrm{Var}\left\{\widetilde{\mathcal{Z}}\right\}$
20	0.985506	0.985520	2.180×10^{-7}	9.495×10^{-8}	5.069×10^{-10}
36	0.909608	0.909620	1.255×10^{-6}	6.692×10^{-7}	2.616×10^{-9}
50	0.667960	0.667936	3.384×10^{-6}	2.452×10^{-7}	1.064×10^{-9}
60	0.456285	0.456283	3.785×10^{-6}	3.130×10^{-8}	3.210×10^{-11}
71	0.372807	0.372810	3.568×10^{-6}	2.015×10^{-8}	5.113×10^{-12}

Table 1: Evolution of variances of $\widehat{\Phi}(\vec{\mathcal{C}})$, $\widetilde{\mathcal{F}}$ and $\widetilde{\mathcal{Z}}$ as a function of the demand d, with $p = 0.9$.

Table 2 compares the performance of $\widehat{\Phi}(\vec{\mathcal{C}})$, $\widetilde{\mathcal{F}}$ and $\widetilde{\mathcal{Z}}$ defined by (1), (7) and (10), respectively. At Column 2, we give the variance-reduction ratio achieved by the estimator $\widetilde{\mathcal{F}}$ when compared to the standard sampling $\widehat{\Phi}(\vec{\mathcal{C}})$. At Column 3, we present the variance-reduction ratio achieved by the estimator $\widetilde{\mathcal{Z}}$ relatively to $\widehat{\Phi}(\vec{\mathcal{C}})$. This ratio is always greater than the variance-reduction ratio achieved by the estimator $\widetilde{\mathcal{F}}$. It results that $\widetilde{\mathcal{Z}}$ is more accurate than $\widetilde{\mathcal{F}}$. At Column 4, the quantity $T(\widehat{\Phi}(\vec{\mathcal{C}}))/T(\widetilde{\mathcal{Z}})$ gives the ratio of the time required to collect 2^{16} trials for $\widehat{\Phi}(\vec{\mathcal{C}})$ and the time required to collect 2^{16} trials for $\widetilde{\mathcal{Z}}$. Because this ratio is always smaller than 1, the reduced variance has been achieved at a higher cost per trial for $\widetilde{\mathcal{Z}}$ than for the standard estimator. The performance parameter that takes into account both variance-reduction ratio and time ratio is the product

$$\frac{W(\widehat{\Phi}(\vec{\mathcal{C}}))}{W(\widetilde{\mathcal{Z}})} = \frac{\mathrm{Var}\left\{\widehat{\Phi}(\vec{\mathcal{C}})\right\}}{\mathrm{Var}\left\{\widetilde{\mathcal{Z}}\right\}} \times \frac{T(\widehat{\Phi}(\vec{\mathcal{C}}))}{T(\widetilde{\mathcal{Z}})}.$$

This parameter is called the *speedup* of the estimator $\widetilde{\mathcal{Z}}$ with respect to $\widehat{\Phi}(\vec{\mathcal{C}})$ and gives the time that the estimator $\widehat{\Phi}(\vec{\mathcal{C}})$ requires to obtain the same variance obtained by the estimator $\widetilde{\mathcal{Z}}$ in one unit of time. The speedup values, presented at Column 5, are always greater than 1, showing a better performance of the estimator $\widetilde{\mathcal{Z}}$.

As the cost of the estimating by $\widetilde{\mathcal{F}}$ is about the same as the cost of estimating by $\widehat{\Phi}(\vec{\mathcal{C}})$ [12], the speedup of $\widetilde{\mathcal{F}}$ with respect to the standard estimator is close to the variance-reduction ratio, and the speedup of $\widetilde{\mathcal{Z}}$ with respect to the stratified sampling estimator can be estimated by the ratio of the values in Columns 5 and 2, for each case. These speedups are tabulated at Column 6. By comparing the speedup values, we deduce that our methodology is more attractive than the stratified sampling in [12].

We can remark from the values $(g_u(U(\Omega)) - g_l(U(\Omega)))$, tabulated at Column 7, that the estimator $\widetilde{\mathcal{F}}$ is as better as the bounds are close, that is the value of $(g_u(U(\Omega)) - g_l(U(\Omega)))$ is close to 0. This effect does not appear for $\widetilde{\mathcal{Z}}$, due to the recursion. However, we can observe that it has the properties of the exact evaluation, that is it is more efficient when the demand is close to the maximal s,t-flow.

d	$\frac{\mathrm{Var}\left\{\widehat{\Phi}(\vec{\mathcal{C}})\right\}}{\mathrm{Var}\left\{\widetilde{\mathcal{F}}\right\}}$	$\frac{\mathrm{Var}\left\{\widehat{\Phi}(\vec{\mathcal{C}})\right\}}{\mathrm{Var}\left\{\widetilde{\mathcal{Z}}\right\}}$	$\frac{T(\widehat{\Phi}(\vec{\mathcal{C}}))}{T(\widetilde{\mathcal{Z}})}$	$\frac{W(\widehat{\Phi}(\vec{\mathcal{C}}))}{W(\widetilde{\mathcal{Z}})}$	$\frac{W(\widetilde{\mathcal{F}})}{W(\widetilde{\mathcal{Z}})}$	$(g_u(U(\Omega)) - g_l(U(\Omega)))$
20	2.30	430	0.156	67	29	0.4686
36	1.87	479	0.215	103	55	0.6126
50	13.80	3180	0.238	758	55	0.3416
60	120.94	117928	0.327	38610	319	0.1296
71	180.11	697790	0.461	321988	1787	0.1050

Table 2: Evolution of the speedup of the estimator $\widetilde{\mathcal{Z}}$ with respect to $\widehat{\Phi}(\vec{\mathcal{C}})$ as a function of the demand d, with $p = 0.9$.

This gain in accuracy of $\widetilde{\mathcal{Z}}$ with respect to $\widetilde{\mathcal{F}}$ is improved in the case of highly reliable components. For example, when the elementary reliabilities are set to 0.99, we obtain the speedups values from 18159 to 2048026.

6 Conclusions

The problem of the exact evaluation of the probability that the maximum st-flow exceeds a fixed value d in a stochastic flow network is an NP-hard problem. Consequently, algorithms to resolve it exactly have a high computational cost. When exact algorithms fail or when their computational time is prohibitive,

Monte Carlo methods can supply an estimate in a reasonable time. In this paper, we have proposed to use the decomposition methodology of Doulliez and Jamoulle to construct a new stratified sampling estimator. We have shown that it belongs to the variance-reduction family and that it is more accurate than the previous stratified sampling estimator based on the same decomposition [12]. We deduce from the results of several tests that our methodology offers substantial gains. Further research can be performed in two directions. On one hand, one can attempt to accomplish the M decompositions such that the bounds are close, at each call of the procedure Bounds(). This will lead to more accurate estimates. On the other hand, one can quantify the effects of varying the number of decompositions M in order to establish some rules to suitably fix this parameter. An other interesting point is to study the incorporation of recursive stratification within the undetermined sets.

References

[1] C. Alexopoulos, A note on state space decomposition methods for analyzing stochastic flow networks, *IEEE Transaction on Reliability*, 1995, 44, p. 354-357.

[2] M.O. Ball, Computational complexity of network reliability analysis an overview, *IEEE Transactions on Reliability*, 1986, 35, p. 230-239.

[3] S. Bulteau and M. El Khadiri. A Recursive Importance Sampling Estimator for a Flow Network Reliability Problem. *submitted to Naval Research Logistics*, 1997.

[4] S. Bulteau, and M. El Khadiri, A Monte Carlo algorithm based on a state space decomposition methodology for flow network reliability evaluation, Technical Report PI 1012, I.R.I.S.A., Campus de Beaulieu, Rennes, France, 1996.

[5] H. Cancela, and M. El Khadiri, A recursive variance-reduction algorithm for estimating communication-network reliability, *IEEE Transaction on Reliability*, 1995, 44, p. 599-602.

[6] J. Carlier, and O. Theologou, Factoring & reductions for networks with imperfect vertices, *IEEE Transaction on Reliability*, 1991, 40, p. 210-217.

[7] P. Doulliez, and E. Jamoulle, Transportation networks with random arc capacities, *R.A.I.R.O.*, 1972, 3, p. 45-59.

[8] M. El Khadiri, Direct evaluation and simulation of communication network reliability parameters Sequential and memory distributed parallel algorithms, PhD thesis, Rennes I, Campus de Beaulieu, 35042 Rennes, France, December 1992.

[9] T. Elperin, I. Gertsbakh, and M. Lomonosov, Estimation of network reliability using graph evolution models, *IEEE Transaction on Reliability*, 1991, 40, p. 572-581.

[10] J.R. Evans, Maximal flow in probabilistic graphs - the discrete case, *Networks*, 1976, 6, p. 161-183.

[11] G.S. Fishman, Principles of Discrete Event Digital Simulation, John Wiley and Sons. Inc., 1978.

[12] G.S. Fishman, and T.D. Shaw, Evaluating reliability of stochastic flow networks, *Probability in the Engineering and Informational Sciences*, 1989, 3, p. 493-509.

[13] J.M. Hammersley, and D.C. Handscomb, Monte Carlo Methods, Halsted Press, Wiley and Sons. Inc., New York, 1979.

[14] S. H. Lee, Reliability in a flow network, *IEEE Transaction on Reliability*, 1980, 29, p. 24-26.

[15] O. Theologou, Contribution to network reliability evaluation, PhD thesis, Dept. of Computer Science, University of Compiègne, Compiègne, France, 1990.

[16] K.S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, Prentice-Hall, Inc., Englewood Cliffs, 1982.



Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399