Proceedings of the 2011 Winter Simulation Conference
S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, eds.

GRAPH REDUCTIONS TO SPEED UP IMPORTANCE SAMPLING-BASED STATIC RELIABILITY ESTIMATION

Pierre L'Ecuyer

DIRO, Université de Montreal
C.P. 6128, Succ. Centre-Ville
Montréal (Québec), H3C 3J7, CANADA

Samira Saggadi
Bruno Tuffin

INRIA Rennes Bretagne Atlantique
Campus Universitaire de Beaulieu
35042 Rennes Cedex, FRANCE

ABSTRACT

We speed up the Monte Carlo simulation of static graph reliability models by adding graph reductions to zero-variance importance sampling (ZVIS) approximation techniques. ZVIS approximation samples the status of links sequentially, and at each step we check if series-parallel reductions can be performed. We present two variants of the algorithm and describe their respective advantages. We show that the method satisfies robustness properties as the reliability of links increases. We illustrate theoretically on small examples and numerically on large ones the gains that can be obtained, both in terms of variance and computational time.

1 INTRODUCTION

Reliability analysis is a critical issue in many areas such as telecommunications, transportation systems, and energy production plants, among others. We focus in this paper on *static models*, where time is not an explicit variable. More specifically, we consider a non-oriented graph $\mathscr{G} = (\mathscr{N}, \mathscr{L})$ where $\mathscr{N}$ is the set of m nodes, and $\mathscr{L} = \{1, \ldots, \ell\}$ is the set of links connecting the nodes. Nodes are assumed perfect, in the sense that they never fail. Our goal is to compute the probability $u(\mathscr{G})$ that a given subset $\mathscr{K}$ of the set of nodes is *not* connected, given that links may fail, with probability q_i for link i ($1 \le i \le \ell$), and failures are assumed to occur *independently* across links.

Formally, if we denote by X_i the (Bernoulli) random variable such that $X_i = 1$ if link $i \in \mathscr{L}$ works (with probability $1 - q_i$) and $X_i = 0$ otherwise, a random state vector is given by

$$X = (X_1, \ldots, X_\ell).$$

We define a function ϕ such that, for each state (or *configuration*) $x = (x_1, \ldots, x_\ell) \in \{0,1\}^\ell$, $\phi(x) = 1$ if the set $\mathscr{K}$ is *not* contained in a single connected component of the graph, when connections are made by links i for which $x_i = 1$, and $\phi(x) = 0$ otherwise. We then have

$$u(\mathscr{G}) = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^\ell} \phi(x) \mathbb{P}[X = x] = \sum_{x \in \{0,1\}^\ell} \phi(x) \prod_{i=1}^{\ell} (q_i(1 - x_i) + (1 - q_i)x_i).$$

Computing this sum requires to evaluate ϕ for 2^ℓ terms. There exist exact combinatorial and bounding techniques that allow to reduce the computational time (Rubino 1998), but computing $u(\mathscr{G})$ is known to be a #P-complete problem in general (Ball and Provan 1982, Colbourn 1987).

Monte Carlo simulation is therefore a relevant method, at least for large graphs, to estimate $u(\mathscr{G})$. Standard Monte Carlo consists in generating n independent copies of X (that is of the random graph), labeled $X^{(1)}, \ldots, X^{(n)}$, and to take as an unbiased estimator of $u(\mathscr{G})$ the average value $(1/n)\sum_{j=1}^{n} \phi(X^{(j)})$.

¿From the Central Limit Theorem, one can obtain a confidence interval with confidence $1-\alpha$ of half-width $c_\alpha \sigma/\sqrt{n}$, where c_α is the $1-\alpha/2$ quantile of the standard normal distribution (with mean 0 and variance 1) and σ is the standard deviation of $\phi(X)$. For standard Monte Carlo, $\phi(X)$ being a Bernoulli random variable, it is easy to check that $\sigma^2 = u(\mathscr{G})(1-u(\mathscr{G}))$.

But reliability analysis usually deals with rare events, i.e., rare cases such that the set of nodes in $\mathscr{K}$ are not connected. This means a low probability $u(\mathscr{G})$, typically 10^{-9} and even less. In that case the Central Limit Theorem cannot be applied except for a very large n, but even if it could, the relative half-width of the confidence interval, $c_\alpha \sigma/(\sqrt{n}u(\mathscr{G}))$ is approximately $c_\alpha/(\sqrt{n}\sqrt{u(\mathscr{G})})$ and increases to infinity as $u(\mathscr{G}) \to 0$. It means that getting a given relative accuracy requires an increasing sample size as the unreliability $u(\mathscr{G})$ decreases. This illustrates the inefficiency of standard Monte Carlo method to deal with rare events. The rarity framework we will consider is when for each $i \in \mathscr{L}$, there are constants $a_i > 0$ and $b_i \geq 0$ independent of ε such that

$$q_i = a_i \varepsilon^{b_i}. \tag{1}$$

It can then be seen that $u(\mathscr{G}) \to 0$ as $\varepsilon \to 0$ (L'Ecuyer et al. 2011). This type of parameterization is inspired from Shahabuddin (1994), who applied it to the evaluation of highly reliable Markovian systems. Thus we typically consider large but fixed topologies for which simulation is required, and look at an asymptotic regime under which individual reliabilities decrease to zero.

There exists an extensive literature on the design of efficient Monte Carlo methods when dealing with rare events; for a general overview, see for example Asmussen and Glynn (2007) and Rubino and Tuffin (2009). The general idea is to design an estimator Y of the probability of interest ($u(\mathscr{G})$, for us) whose relative error $\mathrm{RE}[Y] = (\mathrm{Var}[Y])^{1/2}/\mathbb{E}[Y]$ is "controlled" as $\mathbb{E}[Y] = u(\mathscr{G})$ gets closer to zero, in order to ensure that the sample size required to get a given relative accuracy is also under control. An often considered property of robustness to rarity is *bounded relative error* (BRE) verified by an estimator Y if $\mathrm{RE}[Y]$ remains bounded as $u(\mathscr{G}) \to 0$, so that the sample size needed to get a specified relative accuracy is bounded whatever the rarity of the event. An even better property is *vanishing relative error* (VRE), meaning that $\mathrm{RE}[Y] \to 0$ when $u(\mathscr{G}) \to 0$ (L'Ecuyer et al. 2010). Work-normalized versions of those properties have also been defined in L'Ecuyer et al. (2010) and will be considered here.

Many techniques have been developed specifically for the network reliability problem we are looking at; for a survey, see Cancela, El Khadiri, and Rubino (2009). The most notable ones are the recursive variance reduction technique of Cancela and El Khadiri (2003), which applies recursive conditional Monte Carlo, the turnip method (Gertsbakh and Shpungin 2010), whose basic idea is to replace the static model by a dynamic one (using auxiliary variables) in which each link becomes operational at a random time and the unreliability is estimated by the conditional probability that $\phi(X) = 1$ given the order in which the links become operational, the algorithm of Botev et al. (2011), based on the generalized splitting technique of Botev and Kroese (2010) applied to a model with auxiliary variables, and the approximate zero-variance importance sampling (ZVIS) of L'Ecuyer et al. (2011), summarized in the next section, which we propose to improve. Among those techniques, the turnip and ZVIS are shown to satisfy BRE in general, and ZVIS verifies VRE in several cases. ZVIS was also combined with conditional Monte Carlo in Cancela et al. (2010). Some similitudes can be remarked between ZVIS and Ross (1994). In that paper, the author looks at an estimator based on minimal cuts to reduce the variance and proposes to combine it with IS, but the goal is not to approach the zero-variance importance sampling. Under our asymptotic regime, BRE can be proved for this method though.

Our contribution in this paper is to add graph reductions to the (already efficient) ZVIS algorithm of L'Ecuyer et al. (2011). Under this importance sampling scheme, links are sampled sequentially, the probability that a link fails being dependent on the state of previously sampled links. But fixing the state of a link can lead to some newly possible graph reductions. We consider *series-parallel* reductions defined in Section 3. Those reductions, applied at each step, lead in most cases to variance reductions, but also to computational time reductions even if it requires some work, because fewer links need to be sampled. We describe two implementations, depending on whether the graph reduction is applied after computing the

ZVIS approximation and sampling the link state, or the ZVIS approximation anticipates what would be the reduced graph whatever the link state. We discuss their respective advantages, and illustrate the gains numerically. We also show that, similarly to ZVIS, the combined method satisfy BRE in general. It can satisfy VRE in some cases.

The paper is organized as follows. Section 2 recalls the ZVIS simulation method developed by L'Ecuyer et al. (2011) and that we aim to improve here by using graph reductions. Section 3 describes the graph reduction we use, and discuss how they can be efficiently combined with the IS scheme to get a larger variance reduction than for ZVIS alone, because the unreliability can be better approached, combined with a smaller computational time. The gains are illustrated theoretically on small toy examples and BRE is shown to be satisfied. Section 4 presents the numerical results on a large example to illustrate the power of the method, and we conclude in Section 5.

2 IMPORTANCE SAMPLING-BASED SIMULATION

We recall the method of L'Ecuyer et al. (2011). The idea is to generate the link states $X_1,\ldots,X_\ell$, in that (arbitrary) order. Note that if we generate $X_i = 0$ (for $1 \le i \le \ell$) with probability

$$\tilde{q}_i = \frac{q_i u_{i+1}(x_1,\ldots,x_{i-1},0)}{q_i u_{i+1}(x_1,\ldots,x_{i-1},0) + (1-q_i) u_{i+1}(x_1,\ldots,x_{i-1},1)}, \tag{2}$$

where $u_i(x_1,\ldots,x_{i-1})$ is the unreliability of the graph *given that* $X_1 = x_1,\ldots,X_{i-1} = x_{i-1}$, then the unbiased estimator $\phi(X)L(X)$ has a zero variance, where $L(x) = \prod_{i=1}^{\ell} L_i(x_i)$ and

$$L_i(x_i) \quad = \quad x_i \frac{1-q_i}{1-\widetilde{q}_i} + (1-x_i)\frac{q_i}{\widetilde{q}_i}.$$

In other words, if we sample the link states one after the other, taking into account in the sampling probability the state of previous links according to (2), then the simulation always yields the exact result $u(\mathscr{G})$. In this formulation, we have changed the probability measure, introducing a bias that is adjusted thanks to the *likelihood ratio* $L(X)$; this is an implementation of the *importance sampling* (IS) simulation technique (Asmussen and Glynn 2007).

But sampling according to (2) requires the knowledge of unreliabilities $u_i(x_1,\ldots,x_{i-1})\ \forall i \in \{1,\ldots,\ell\}$, and therefore the value $u(\mathscr{G}) = q_1 u_1(0) + (1-q_1)u_1(1)$ that we are trying to estimate. If we knew those values, there would be no need for simulation. The principle is therefore to choose a simple-to-compute approximation $\widehat{u}_{i+1}(x_1,\ldots,x_i)$ of $u_{i+1}(x_1,\ldots,x_i)\ \forall i$ that we will plug into (2) in place of the $u_{i+1}(x_1,\ldots,x_i)$. The approximation we use is called *mincut-maxprob approximation*. To define it formally, we need some definitions and notations. For each i, given $x_1,\ldots,x_i$ (assumed fixed), note by $\mathscr{G}_i = \mathscr{G}_i(x_1,\ldots,x_i)$ the graph obtained from $\mathscr{G}$ by removing all links $j \le i$ for which $x_j = 0$ and forcing the links j such that $x_j = 1$ to be operational. A *cut* (or $\mathscr{K}$-cut) in the graph $\mathscr{G}_i$ is a set of links (containing no link $j \le i$ for which $x_j = 1$) such that if we remove them from $\mathscr{G}_i$, not all the nodes in $\mathscr{K}$ are in the same connected component of the resulting graph. A *mincut* (minimal cut) of $\mathscr{G}_i$ is a cut that contains no other cut than itself. Let $\mathscr{C}_i$ be the set of mincuts in $\mathscr{G}_i$. The probability of a cut $\gamma \in \mathscr{C}_i$ is defined as the probability that all links of the cut fail. Let γ_i be a mincut of maximal probability in $\mathscr{C}_i$. We use this probability as our approximation $\widehat{u}_{i+1}(x_1,\ldots,x_i)$ of $u_{i+1}(x_1,\ldots,x_i)$ at step i, and the IS is called the zero-variance importance sampling (ZVIS) approximation scheme.

It is shown in L'Ecuyer et al. (2011) that when rarity comes from rare failure of individual links, i.e., under (1), then the ZVIS estimator verifies BRE in general as $\varepsilon \to 0$, and can also satisfy VRE under some additional conditions, satisfied by several network topologies. The intuition is that, thanks to our approximation, and under our ε parameterization, the mincut-maxprob approximation $\widehat{u}_{i+1}(x_1,\ldots,x_i)$ and $u_{i+1}(x_1,\ldots,x_i)$ are of the same order in ε, meaning that our approximation is asymptotically valid.

3 GRAPH REDUCTIONS TO DECREASE THE WORK-NORMALIZED VARIANCE

Each time a link state is generated by the ZVIS algorithm, the graph evolves according to these rules: at step i ($1 \leq i \leq \ell$),

- either $X_i = 0$ which means that the link is removed,
- or $X_1 = 1$ which means that the link is fixed, and can then be removed by merging the two nodes it links.

In each case, the graph topology is simplified (and modified). At each step i, we can therefore search if graph reductions can be applied, in order to simplify the topology, and potentially gain in terms of variance (by using a conditional expectation) and computational time (because the size of the graph is smaller). This will be illustrated afterwards. Two types of graph reductions are investigated:

- *Series reduction*: Assume that node $s \in \mathscr{N}$ has only two incident links, l_1 and l_2, connecting it to nodes s_1 and s_2 respectively. If $s \notin \mathscr{K}$, node s can be removed and links l_1 and l_2 merged into a single one, with unreliability $q = 1-(1-q_{l_1})(1-q_{l_2})$. Remark that the case $s \in \mathscr{K}$ can hardly be treated without further topology information.
- *Parallel reduction*: if there are two (parallel) links l_1 and l_2 both connecting nodes s_1 and s_2, those two links can be merged into a single one, with unreliability $q = q_{l_1} q_{l_2}$.

We consider two different combinations of the reductions with the IS procedure. The first approach is called the *posterior reduction* (PR) because it applies the reductions after the link state is sampled, while the second is called the *look-ahead reduction* (LAR) because it anticipates what would be the reductions whatever the link state, when computing the mincut-maxprob approximations. Of course, we also test if the initial graph can be reduced. Each time we want to sample the state of a link i:

- for the PR, this link will be considered failed with probability

$$\hat{q}_i^{(1)} = \frac{q_i \hat{u}_{i+1}(\mathscr{G}_i^r, 0)}{q_i \hat{u}_{i+1}(\mathscr{G}_i^r, 0) + (1-q_i)\hat{u}_{i+1}(\mathscr{G}_i^r, 1)}, \tag{3}$$

  where $\mathscr{G}_i^r$ is the graph resulting from previous link samplings and reductions, and $\hat{u}_{i+1}(\mathscr{G}_i^r, 0)$ (respectively $\hat{u}_{i+1}(\mathscr{G}_i^r, 1)$) is the mincut-maxprob approximation of $\mathscr{G}_i^r$ with link i state X_i fixed to 0 (respectively 1). After the state of link i is determined, the link i is removed if $X_i = 0$ and compressed if $X_i = 1$ (by merging the two nodes it joins, and again removing the link), we can search for new reductions, leading to a new graph $\mathscr{G}_{i+1}^r$. We therefore end up with a direct application of the zero-variance IS algorithm of L'Ecuyer et al. (2011), but applied to (potentially) successively reduced graphs.
- For the LAR, we rather use as the probability that i is failed:

$$\hat{q}_i^{(2)} = \frac{q_i \hat{u}_{i+1}(\mathscr{G}_{i,0}^r)}{q_i \hat{u}_{i+1}(\mathscr{G}_{i,0}^r) + (1-q_i)\hat{u}_{i+1}(\mathscr{G}_{i,1}^r)}, \tag{4}$$

  where $\mathscr{G}_{i,k}^r$ for $k \in \{0,1\}$ is the graph *reduced after* setting $X_i = k$, and $\hat{u}_{i+1}(\mathscr{G}_{i,k}^r)$ the corresponding mincut-maxprob approximation. This requires to make two copies of the graph, setting $X_i = 0$ for the first and $X_i = 1$ for the other, the two resulting graphs being reduced according to the above rules, if possible. With respect to the first implementation, we therefore anticipate what would be the reductions depending on the value of X_i. This is a type of *one step look-ahead* strategy. Afterwards, when link i's state is effectively sampled, we choose the appropriate already reduced graph.

The gains we expect to obtain with those reductions come from different aspects:

- *Gain in terms of computational time*: Reducing the graph means spending some time to search for reductions, and to make copies of the graph. But it allows to decrease the number of links to sample, and therefore the number of steps in the ZVIS algorithm and the number of mincut-maxprob approximations that need to be computed, which may lead to substantial computational savings. This is illustrated in Example 1 (below), for which cascading reductions occur.
- *Gain in terms of variance*: applying reductions allows to obtain a better mincut-maxprob approximation of the graph unreliabilities at the different steps, usually resulting in smaller variance. On the other hand, the graph obtained after a reduction is in general different from a graph that would be obtained without reduction (depending on the link ordering), and consequently it is not possible to ensure that variance will be reduced, even though it usually is reduced.
- Comparing the two implementations, the LAR algorithm requires some additional time to make copies of the graph and to perform twice more reductions at any given step. On the other hand, computing the mincut-maxprob on an already reduced graph takes a shorter time than before proceeding to a reduction. Moreover we usually get a better approximation of the zero-variance IS with this procedure (see Example 1 below), which reduces the variance.

Potential gains, and the difference between the two implementations, are illustrated on the following toy example.

Example 1 Consider the graph of Figure 1, where we want to compute the probability that the gray nodes A and D are disconnected. Links are assumed homogeneous, with unreliability $q_i = \varepsilon$ for $i = 1,\dots,5$. The graph unreliability is $u(\mathscr{G}) = 2\varepsilon^2 + 2\varepsilon^3 - 5\varepsilon^4 + 2\varepsilon^5$. For this example, it is shown in L'Ecuyer et al. (2011) that the zero-variance IS approximation already satisfies VRE. Note that none of the reductions described above can be applied to this initial graph.

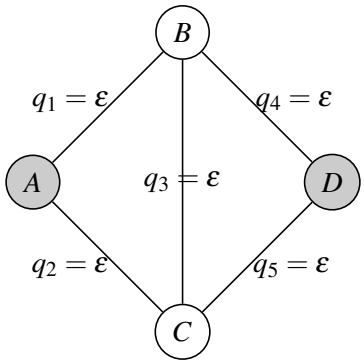

Figure 1: Graph topology with five links and two nodes requiring to be connected.

- With our first implementation, we first sample link 1. In this case the mincut-maxprob approximations are $\hat{u}_2(\mathscr{G},0) = \varepsilon$ and $\hat{u}_2(\mathscr{G},1) = \varepsilon^2$. This gives an IS probability $\varepsilon^2/(\varepsilon^2 + (1-\varepsilon)\varepsilon^2) = 1/(2-\varepsilon)$ to have $X_i = 0$.
  - If $X_1 = 1$, the likelihood ratio when sampling this link is $(2-\varepsilon)$. The graph can then be reduced by compressing link 1, merging nodes A and B, then a parallel reduction of links 2 and 3 can be applied, leading to a link with unreliability ε^2. This new link is then in series with link 5, leading to a reduction to a link with unreliability $1-(1-\varepsilon)(1-\varepsilon^2) = \varepsilon + \varepsilon^2 - \varepsilon^3$. The resulting graph is then just made of two parallel links between nodes: the one just determined and link 4, which can therefore be reduced to a graph with a single link with unreliability $\varepsilon^2 + \varepsilon^3 - \varepsilon^4$. In this case, by IS, the link is necessarily considered failed (because $\hat{u}_3(\mathscr{G}_2^r,1) = 0$), and the

likelihood is $\varepsilon^2+\varepsilon^3-\varepsilon^4$. This gives, with probability $(1-\varepsilon)/(2-\varepsilon)$, the estimate

$$(2-\varepsilon)(\varepsilon^2+\varepsilon^3-\varepsilon^4).$$

– If $X_1=0$ (with probability $1/(2-\varepsilon)$ under IS, leading to a likelihood ratio when sampling this link $\varepsilon(2-\varepsilon)$), link 1 is removed. Links 3 and 4 are then in series and can be reduced, leading to a new link of unreliability $1-(1-\varepsilon)^2=2\varepsilon-\varepsilon^2$ which becomes a parallel link with link 5, reduced then to become a link with unreliability $2\varepsilon^2-\varepsilon^3$. But we then end up with two links in series, the one just described, and link 2, which can be reduced to lead to a single link with unreliability $1-(1-\varepsilon)(1-2\varepsilon^2+\varepsilon^3)=\varepsilon+2\varepsilon^2-3\varepsilon^3+\varepsilon^4$. Here again, under IS this link is sampled as failed with probability 1, leading to the estimate

$$\varepsilon(2-\varepsilon)(\varepsilon+2\varepsilon^2-3\varepsilon^3+\varepsilon^4)$$

with probability $1/(2-\varepsilon)$.

The second moment of the estimator, that we note $\mathbb{E}[(Y_{\rm PR})^2]$, is then

$$\mathbb{E}[(Y_{\rm PR})^2]=(\varepsilon(2-\varepsilon)(\varepsilon+2\varepsilon^2-3\varepsilon^3+\varepsilon^4))^2\frac{1}{2-\varepsilon}+((2-\varepsilon)(\varepsilon^2+\varepsilon^3-\varepsilon^4))^2\frac{1-\varepsilon}{2-\varepsilon}=4\varepsilon^4+o(\varepsilon^4),$$

and, computing the relative variance, we get

$$\frac{\mathbb{E}[(Y_{\rm PR})^2]-(u(\mathscr{G}))^2}{(u(\mathscr{G}))^2}\to 0 \text{ as } \varepsilon\to 0,$$

meaning that VRE is satisfied. Obviously, the successive reductions have diminished the number of links to be sampled and the number of steps in the ZVIS algorithm, resulting in a time saving that (usually) outweighs the additional time spent on applying the reductions.

- With the LAR implementation, the mincut-maxprob approximations are computed *anticipating* what would be the reductions when $X_i=0$ or 1. In that case, the reductions are the two just described in the case of the first method. This yields $\hat{u}_2(\mathscr{G}_{1,0})=\varepsilon+2\varepsilon^2-3\varepsilon^3+\varepsilon^4$ and $\hat{u}_2(\mathscr{G}_{1,1})=\varepsilon^2+\varepsilon^3-\varepsilon^4$. The probability that $X_1=0$ is then

$$\hat{q}_1^{(2)}=\frac{\varepsilon(\varepsilon+2\varepsilon^2-3\varepsilon^3+\varepsilon^4)}{\varepsilon(\varepsilon+2\varepsilon^2-3\varepsilon^3+\varepsilon^4)+(1-\varepsilon)(\varepsilon^2+\varepsilon^3-\varepsilon^4)}.$$

– If $X_1=1$, the likelihood ratio when sampling this link is $(1-\varepsilon)/(1-\hat{q}_1^{(2)})$. The reduced graph is the one with a single link, with unreliability $\hat{u}_2(\mathscr{G}_{1,1})$. This reduced link is considered failed under IS with probability 1 (so the likelihood ratio is $\hat{u}_2(\mathscr{G}_{1,1})$), leading to the estimate

$$\frac{1-\varepsilon}{1-\hat{q}_1^{(2)}}\hat{u}_2(\mathscr{G}_{1,1})=(1-\varepsilon)\hat{u}_2(\mathscr{G}_{1,1})+\varepsilon\hat{u}_2(\mathscr{G}_{1,0})=u(\mathscr{G}).$$

– If $X_1=0$, the likelihood ratio when sampling this link is $\varepsilon/\hat{q}_1^{(2)}$. The reduced graph is the one with a single link, with unreliability $\hat{u}_2(\mathscr{G}_{1,0})$. This reduced link is considered failed under IS with probability 1 (so the likelihood ratio is $\hat{u}_2(\mathscr{G}_{1,0})$), leading to the estimate

$$\frac{\varepsilon}{\hat{q}_1^{(2)}}\hat{u}_2(\mathscr{G}_{1,0})=(1-\varepsilon)\hat{u}_2(\mathscr{G}_{1,1})+\varepsilon\hat{u}_2(\mathscr{G}_{1,0})=u(\mathscr{G}).$$

We therefore end up with an estimator always giving the exact result $u(\mathscr{G})$, that is, a zero-variance estimator. Here too, the number of sampled links and ZVIS iterations are significantly reduced. An important remark is that the approximations $\hat{u}_{i+1}(\mathscr{G}^r_{i,k})$ in the LAR algorithm are better approximations of the searched graph unreliabilites than the $\hat{u}_{i+1}(\mathscr{G}^r_i,k)$ in the PR algorithm. Indeed, with LAR, reductions are applied before the approximations are computed, and the probability of a mincut can correspond to the probability of a *set* of cuts with PR because cuts may be "grouped". This is what happens in our example when sampling the first link: for the LAR algorithm, the mincut-maxprob approximations of the reduced graph are actually the exact unreliabilities of the graph conditionally to the value of X_1, which is not the case for the PR algorithm.

With full generality, the algorithms have the following robustness properties, as $\varepsilon \to 0$.

Proposition 1 Our algorithms satisfy BRE.

Proof. The proof follows exactly the arguments of Theorem 2 (and Theorem 4) in L'Ecuyer et al. (2011). It is indeed straightforward to check as in L'Ecuyer et al. (2011) that, for $k \in \{0,1\}$ and $\forall i$, $\hat{u}_{i+1}(\mathscr{G}^r_i,k) = \Theta(u_{i+1}(\mathscr{G}^r_i,k))$ and $\hat{u}_{i+1}(\mathscr{G}^r_{i,k}) = \Theta(u_{i+1}(\mathscr{G}^r_{i,k}))$. □

Similarly to L'Ecuyer et al. (2011), VRE can also be verified under some conditions. This is actually illustrated *empirically* on the numerical example below.

4 A NUMERICAL ILLUSTRATION

As a numerical illustration, we consider a graph often used as a benchmark for network reliability evaluation techniques (Cancela et al. 2009). It is described in Figure 2, and made of 20 nodes and 30 links, with the dodecahedron topology as shown. We consider the homogeneous case, where all links have the same

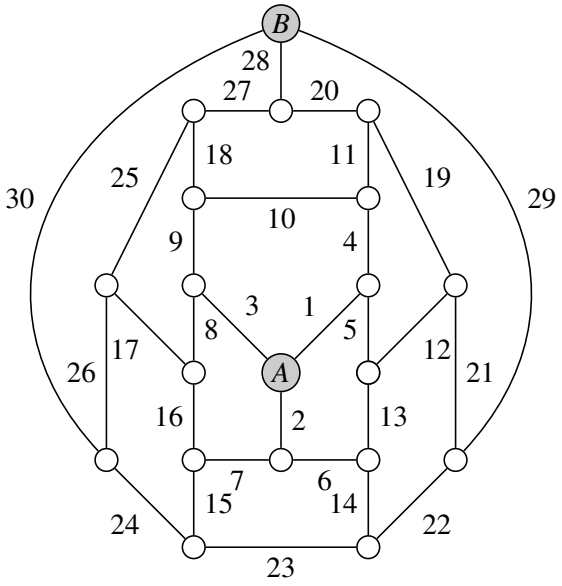

Figure 2: A dodecahedron graph with 20 nodes and 30 links.

unreliability ε, and we want to compute the probability that nodes A and B are disconnected. Links are ordered somewhat arbitrarily, according to their numbering in the figure.

Table 1 shows the results obtained for the ZVIS, PR, and LAR algorithms from $n = 10^4$ independent simulation runs and for three different values of ε. We see that the relative error in the fifth column decreases with ε, which indicates (empirically) that VRE is verified for the three methods. The variance, in

Table 1: Empirical results for the ZVIS, PR and LAR algorithms on the dodecahedron topology, obtained from $n = 10^4$ independent simulation runs and three values of ε.

Method	ε	Estimate	Variance	Rel. Err.	Time
ZVIS	10^{-1}	2.8328×10^{-3}	1.1048×10^{-5}	1.1733	15.18
ZVIS	10^{-2}	2.0677×10^{-6}	1.1670×10^{-13}	0.1652	14.35
ZVIS	10^{-3}	2.0074×10^{-9}	1.2714×10^{-20}	0.0561	14.88
PR	10^{-1}	2.8751×10^{-3}	5.5452×10^{-6}	0.8190	12.14
PR	10^{-2}	2.0651×10^{-6}	9.8889×10^{-14}	0.1522	15.33
PR	10^{-3}	2.0068×10^{-9}	9.5548×10^{-21}	0.0487	13.87
LAR	10^{-1}	2.8778×10^{-3}	3.9203×10^{-6}	0.6880	10.29
LAR	10^{-2}	2.0612×10^{-6}	4.4955×10^{-14}	0.1028	7.48
LAR	10^{-3}	2.0051×10^{-9}	2.4094×10^{-21}	0.0244	7.55

the fourth column, is smaller when we use the two series-parallel graph reductions, via either PR or LAR, compared with ZVIS alone. This is due to better mincut-maxprob approximations of the unreliabilities, thanks to the reductions. The gain is larger for LAR than for PR, as was the case in Example 1. Computing times (in seconds) are displayed in the sixth column. We find that despite the work required to clone the graphs and search for reductions, both PR and LAR reduce the overall computing times, presumably due to a decrease of the number of steps in the ZVIS algorithm (fewer links to be sampled). The average number of series-parallel reductions per simulation run were actually about 11.1 for PR and 11.4 for LAR, for the three values of ε. That is, the number of links that need to be sampled in the ZVIS method is reduced from 30 to about 19, with either PR or LAR. The small difference between the two methods is probably due to some configurations with more reductions are favored under LAR algorithm with respect to PR, on this example. We also see from the last column of Table 1 that LAR is faster than PR, despite the additional number of reductions performed in the look-ahead strategy, because the time to compute the mincut-maxprob approximations is smaller on the look-ahead (reduced) graphs. Overall, by applying PR or LAR, we gain both in terms of variance and computing time, hence a reduced work-normalized variance (product of variance and time), with an advantage for LAR.

5 CONCLUSION

We have described how graph reductions can be combined with IS for a static graph reliability estimation problem. We considered series and parallel link reductions. Two algorithms have been defined to get an IS probability closer to the zero-variance one: One for which the reductions are applied after each link sampling (PR), and the other one that anticipates what would be the reductions after the link state determination (LAR). We have illustrated, theoretically on a small example and numerically on a larger one, the gains that can be obtained, both in terms of variance and computational time. The additional work to perform the reductions is more than compensated by the time savings from reducing the number of links to be sampled (and thus the number of steps) in the ZVIS algorithm. We have also shown that BRE is satisfied in general, and VRE in some cases, when the reliability of individual links goes to one.

As next step, we would like to include other types of reductions. For instance if there is a node $s \in \mathcal{N}$ of degree 1, i.e., if there is only one link connected to node s, then there are two possibilities:

- if $s \notin \mathcal{K}$, the state of the link attached to s has no influence on the graph unreliability, therefore, it can be removed from $\mathcal{L}$, and s removed from $\mathcal{N}$;
- if $s \in \mathcal{K}$, call $\mathcal{G}'$ the considered graph, q the unreliability of the link attached to s, and $\tilde{\mathcal{G}}'$ the reduced graph where node s is merged to its neighbor (the resulting node being still considered in $\mathcal{K}$). By conditioning, it is easy to check that $u(\mathcal{G}') = 1 - (1-q)(1-u(\tilde{\mathcal{G}}'))$. We can therefore

merge again node s to its neighbor and focus on the estimation of $u(\tilde{\mathscr{G}}')$. In that case, if we denote by Y the random variable whose expectation is $u(\mathscr{G})$, and L the likelihood ratio when combining IS and reductions up to reaching a reduction of a degree-1 node i in $\mathscr{K}$, with $\mathscr{G}'$ the resulting reduced graph. We then have

$$Y = L(1-(1-q_i)(1-Y')).$$

We can proceed that way recursively, applying the algorithm for $\mathscr{G}'$, up to exhausting the links or ensuring that the nodes in $\mathscr{K}$ are disconnected.

ACKNOWLEDGMENTS

This work has been supported by INRIA's associated team MOCQUASIN to all authors, and NSERC-Canada and a Canada Research Chair to the first author.

REFERENCES

Asmussen, S., and P. W. Glynn. 2007. *Stochastic simulation*. New York: Springer-Verlag.

Ball, M. O., and J. S. Provan. 1982. "Bounds on the Reliability Polynomial for Shellable Independence Systems". *SIAM Journal on Algebraic and Discrete Methods* 3:166–181.

Botev, Z. I., and D. P. Kroese. 2010. "Efficient Monte Carlo Simulation via the Generalized Splitting Method". *Statistics and Computing*. to appear.

Botev, Z. I., P. L'Ecuyer, G. Rubino, R. Simard, and B. Tuffin. 2011. "Static Network Reliability Estimation Via Generalized Splitting". submitted.

Cancela, H., and M. El Khadiri. 2003. "On the RVR simulation algorithm for network reliability evaluation". *IEEE Transactions on Reliability* 52 (2): 207–212.

Cancela, H., M. El Khadiri, and G. Rubino. 2009. "Rare Event Analysis by Monte Carlo Techniques in Static Models". In *Rare Event Simulation Using Monte Carlo Methods*, Edited by G. Rubino and B. Tuffin, 145–170. Wiley. Chapter 7.

Cancela, H., P. L'Ecuyer, G. Rubino, and B. Tuffin. 2010. "Combination of Conditional Monte Carlo and Approximate Zero-Variance Importance Sampling for Network Reliability Estimation". In *Proceedings of the 2010 Winter Simulation Conference*, Edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 1263–1274. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Colbourn, C. J. 1987. *The combinatorics of network reliability*. New York: Oxford University Press.

Gertsbakh, I. B., and Y. Shpungin. 2010. *Models of network reliability*. Boca Raton, FL: CRC Press.

L'Ecuyer, P., J. H. Blanchet, B. Tuffin, and P. W. Glynn. 2010. "Asymptotic Robustness of Estimators in Rare-Event Simulation". *ACM Transactions on Modeling and Computer Simulation* 20 (1): Article 6.

L'Ecuyer, P., G. Rubino, S. Saggadi, and B. Tuffin. 2011. "Approximate Zero-Variance Importance Sampling for Static Network Reliability Estimation". *IEEE Transactions on Reliability*. to appear.

Ross, K. W. 1994. "A New Simulation Estimator of System Reliability". *Journal of the Applied Mathematics and Stochastic Analysis* 7 (3).

Rubino, G. 1998. "Network Reliability Evaluation". In *The state-of-the art in performance modeling and simulation*, Edited by K. Bagchi and J. Walrand. Gordon and Breach Books. Chapter 11.

Rubino, G., and B. Tuffin. (Eds.) 2009. *Rare event simulation using Monte Carlo methods*. Wiley.

Shahabuddin, P. 1994. "Importance Sampling for the Simulation of Highly Reliable Markovian Systems". *Management Science* 40 (3): 333–352.

AUTHOR BIOGRAPHIES

PIERRE L'ECUYER is Professor in the Département d'Informatique et de Recherche Opérationnelle, at the Université de Montréal, Canada. He holds the Canada Research Chair in Stochastic Simulation and Optimization. He is a member of the CIRRELT and GERAD research centers. His main research interests

are random number generation, quasi-Monte Carlo methods, efficiency improvement via variance reduction, sensitivity analysis and optimization of discrete-event stochastic systems, and discrete-event simulation in general. He is currently Editor-in-Chief for *ACM Transactions on Modeling and Computer Simulation*, and Associate/Area Editor for *ACM Transactions on Mathematical Software*, *Statistics and Computing*, *International Transactions in Operational Research*, and *Cryptography and Communications*. More information and his recent research articles are available on-line at http://www.iro.umontreal.ca/∼lecuyer.

SAMIRA SAGGADI is a PhD student at INRIA, Rennes, France. Her research interests are the design of new rare event simulation techniques, and their application to dependability and/or telecommunication problems. Her e-mail is samira.saggadi@inria.fr.

BRUNO TUFFIN received his PhD degree in applied mathematics from the University of Rennes 1 (France) in 1997. Since then, he has been with INRIA in Rennes. He spent eight months as a postdoc at Duke University in 1999. His research interests include developing Monte Carlo and quasi-Monte Carlo simulation techniques for the performance evaluation of telecommunication systems, and developing new Internet-pricing schemes and telecommunication-related economical models. He is currently Associate Editor for *INFORMS Journal on Computing*, *ACM Transactions on Modeling and Computer Simulation* and *Mathematical Methods of Operations Research*. He has written or co-written two books devoted to simulation: *Rare event simulation using Monte Carlo methods* published by John Wiley & Sons in 2009, and *La simulation de Monte Carlo* (in French), published by Hermes Editions in 2010. His email address is bruno.tuffin@inria.fr.