

Approximate zero-variance importance sampling for static network reliability estimation with node failures and application to rail systems

Ajit Rai, Rene Valenzuela, Bruno Tuffin, Gerardo Rubino, Pierre Dersin

► To cite this version:

Ajit Rai, Rene Valenzuela, Bruno Tuffin, Gerardo Rubino, Pierre Dersin. Approximate zero-variance importance sampling for static network reliability estimation with node failures and application to rail systems. Winter Simulation Conference, Dec 2016, Arlington, United States. hal-01398921

HAL Id: hal-01398921

<https://hal.inria.fr/hal-01398921>

Submitted on 18 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**APPROXIMATE ZERO-VARIANCE IMPORTANCE SAMPLING FOR
STATIC NETWORK RELIABILITY ESTIMATION WITH NODE
FAILURES AND APPLICATION TO RAIL SYSTEMS**

Ajit Rai

ALSTOM and INRIA
Campus de Beaulieu, 263 Avenue Général Leclerc
35042 Rennes, FRANCE

Rene C. Valenzuela

RAM Center of Excellence, ALSTOM TIS
Calle de Martínez Villergas, 49
28027 Madrid, SPAIN

Bruno Tuffin
Gerardo Rubino

INRIA
Campus de Beaulieu, 263 Avenue Général Leclerc
35042 Rennes, FRANCE

Pierre Dersin

ALSTOM TIS
48 Rue Albert Dhalenne
93482 St-Ouen, FRANCE

ABSTRACT

To accurately estimate the reliability of highly reliable rail systems and comply with contractual obligations, rail system suppliers such as ALSTOM require efficient reliability estimation techniques. Standard Monte-Carlo methods in their crude form are inefficient in estimating static network reliability of highly reliable systems. Importance Sampling techniques are an advanced class of variance reduction techniques used for rare-event analysis. In static network reliability estimation, the graph models often deal with failing links. In this paper, we propose an adaptation of an approximate Zero-Variance Importance Sampling method to evaluate the reliability of real transport systems where nodes are the failing components. This is more representative of railway telecommunication system behavior. Robustness measures of the accuracy of the estimates, *bounded* or *vanishing relative error* properties, are discussed and results from a real network (Data Communication System used in automated train control system) showing bounded relative error property, are presented.

1 INTRODUCTION

From a reliability point of view, passenger rail systems are large scale systems made up of heterogeneous highly reliable components. Redundancy at different hierarchical levels is used to ensure achievement of system availability requirements (Dersin and Valenzuela 2012). One of the main functions necessary for nominal operation is the communication of different signals between centrally localized computers and trackside/ onboard equipment. In Alstom's urban metro solution the subsystem whose objective is to perform this communication function is called the *data communication system* (DCS). The DCS uses a dual-ring topology to communicate equipment located in different stations (or the track) with centrally located computers. It is configured so that all end-communication equipment has a preferred ring through which it sends its messages (but is able to use the other ring if needed) and all messages are simultaneously transmitted on each ring separately. The availability of the DCS is the probability that all messages between all end-communication devices are successfully transmitted. The model is illustrated in Section 2.

The current approach to predict the availability of such a system involves the creation of a Markov model which characterizes the different failure paths of the network. Typically up to third-order failure paths are included. The selection of which paths to model is made by reliability modeling experts. Resulting models are hard to validate both by other experts and the end-user. Furthermore, it is not clear whether there exist relevant failure paths that have not been modeled. Modeling the communication network as a graph with communication equipment as nodes and communication paths as links overcomes both shortcomings: first, the model can be easily validated by the design expert and the client; and second, by defining successful communication as the existence of a path between the communicating devices, no modeling of failure paths is needed because path finding algorithms can be used to establish connectivity.

The static network reliability problem deals with the estimation of the probability that a given set of nodes in a graph model are connected when each individual component (link or node) is in an UP/ DOWN (working/ failed) state according to their respective probabilities. The case where links are the failing elements is essential in many applications and has been extensively studied (Cancela, Khadiri, and Rubino 2009). However, there is a wide range of applications where nodes are the failing components such as the DCS, e.g., models of network survivability (Gertsbakh, Shpungin, and Vaisman 2014). This requires an adaptation of the existing methods to the case of node failures. Formally, a node failure means that the node becomes nonfunctional and its associated links useless. In the *2-terminal* or *source-to-terminal* reliability problem, two nodes of the graph are fixed and the reliability of the network is defined as the probability of having a path between those two nodes. In such analysis, a node failure causes a higher number of *s-t* paths to become nonfunctional as compared to a link failure (depending on the node's degree). Thus, the reliability of a network would be affected more severely in the case of node failures.

Computing the *unreliability* of highly reliable systems (e.g., the DCS) requires efficient simulation techniques. For large graphs, an exact computation of the unreliability u becomes a NP-hard problem that is impractical to be solved analytically (L'Ecuyer et al. 2011). Monte Carlo methods can estimate u in its crude form (CMC) sampling n stochastically independent realizations of the graph and computing the proportion of these n realizations for which the *s-t* are not connected (L'Ecuyer et al. 2011). For rare events, when $u \ll 1$, the accuracy of the simulation process is captured by the *relative error* RE (ratio of standard deviation and mean value) which is inversely proportional to the event probability u and the number of realizations n (L'Ecuyer et al. 2011, Rubino and Tuffin 2009). Thus, as $u \rightarrow 0$, for a fixed RE, we need excessively large values of n . This increases the computational effort and the cost.

Importance Sampling (IS) is an advanced class of variance reduction techniques for rare-event estimation problems based on the change of the sampling probabilities of the components (i.e., nodes in our case) so that the system failure occurs more frequently. The biasedness of the estimator is removed by multiplying the original estimator with an appropriate likelihood ratio (ratio of the original probability and the new sampling probability) and the estimator is the average over n (Rubino and Tuffin 2009).

This is the general basic framework of IS method. Finding this change of measure is the main difficulty in IS, because if the sampling probabilities which lead to frequent failure are not properly selected, the likelihood ratio may have a huge variance resulting in a bad estimation, even if the failure event is not rare anymore (L'Ecuyer, Mandjes, and Tuffin 2009). The robustness of the estimators in such cases is based on the behavior of the RE properties, like *bounded relative error* (BRE) or *vanishing relative error* (VRE), as described by L'Ecuyer et al. (2011). If the relative width of the confidence interval (CI) on u based on the central-limit theorem (CLT) for a fixed n steadies when $u \rightarrow 0$, the BRE property is valid, and VRE is valid if it tends to zero (L'Ecuyer et al. 2011, L'Ecuyer et al. 2010).

The aim of this paper is to propose and adapt the dynamic importance sampling method based on Monte Carlo simulations as described by L'Ecuyer et al. (2011), considering node failures, and to prove its application on an existing example of a communication network (DCS). We propose an approximation of the zero-variance IS method based on minimal cuts having relatively high failure probability in the subgraph that remains after removing the nonfunctional nodes and their associated links (irrespective of being functional or not, if one of the associated node is failed), while enforcing the states of the nodes

which are functional, at each step of a Markov chain (L'Ecuyer et al. 2011). These cuts approximate the u conditional on the current state, at each step. The networks are analysed as a graph model and the Ford-Fulkerson *maxflow-mincut* algorithm (Sedgewick and Schidlowsky 2003) is adapted for considering flow through nodes. Our estimators show BRE property in general as node reliability increases, and VRE property under additional conditions as proved by L'Ecuyer et al. (2011) for link failure case. The usefulness of the proposed scheme is proved using a quantified measure of work normalized variance (var_{wm}).

The paper is organized as follows: in Section 2, the mathematical model for considering node failures is explained. In Section 3, the inefficiency of crude Monte Carlo (CMC) methods is explained with respect to rare event analysis. In Section 4, the basic idea of Importance Sampling (IS), approximate Zero-Variance IS method and the BRE & VRE robustness properties of estimators are described. In Section 5, we explain the adaptation of the Ford-Fulkerson maxflow-mincut algorithm and the approximate zero-variance IS method based on mincuts for considering node failures is explained. The analysis of the method on various networks, including a case study on existing network of DCS and its results showing BRE or VRE properties are illustrated. Conclusions of the whole study are drawn in Section 6.

2 MATHEMATICAL MODEL

Let us consider an undirected connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ where $\mathcal{N} = \{1, \dots, m\}$ is the set of nodes, and \mathcal{L} is the set of links. The model is static, that is, time is not considered. Links are assumed to always work, but the nodes are subject to (independent) failures. Node $i \in \mathcal{N}$ fails with a probability q_i , where $0 < q_i < 1$. A configuration of the graph (Rubino 1998) is given by the random vector $X = (X_1, \dots, X_m)$, where for all $i \in \mathcal{N}$, $X_i = 1$ or 0 representing the working or failed state of a node i , respectively. Retaining only the functional nodes \mathcal{N}' , we obtain a random partial graph $\mathcal{G}' = (\mathcal{N}', \mathcal{L})$ of \mathcal{G} .

To estimate the probability u that two nodes named s and t (for source and terminal respectively) are not connected in the random graph \mathcal{G} , we define the structure function $\psi(X)$ equal to 1 if s and t are not connected in \mathcal{G}' (or equivalently the configuration X), else as 0 (L'Ecuyer et al. 2011). The expectation $u = \mathbb{E}[\psi(X)]$ or the s - t unreliability is given by (L'Ecuyer et al. 2011):

$$u = \mathbb{E}[\psi(X)] = \sum_{x \in \{0,1\}^m} \psi(x) \mathbb{P}[X = x] = \sum_{x \in \{0,1\}^m} \psi(x) \prod_{i=1}^m (q_i(1-x_i) + (1-q_i)x_i)$$

where $x = (x_1, \dots, x_m)$, \mathbb{P} is the original probability law of the network and \mathbb{E} is the expectation under \mathbb{P} .

The state space having 2^m possible configurations will require an exponentially increasing time to calculate the u from the above formula. The exact evaluation is a NP-hard problem in general (Ball 1986), so approximation techniques like Monte Carlo (MC) are required in such cases. We study the performance of our methodology by parameterizing q_i (under the condition $q_i \rightarrow 0$) as a polynomial function of a rarity parameter $\varepsilon \ll 1$. As explained in L'Ecuyer et al. (2011), for each $i \in \mathcal{N}$, there are independent constants $a_i > 0$ and $b_i \geq 0$ such that $q_i = a_i \varepsilon^{b_i}$. The overall u is a finite sum of products of such possibilities. It is then a polynomial in ε and

$$u = u(\varepsilon) = \Theta(\varepsilon^c) \tag{1}$$

for a constant $c \geq 0$ and Θ is a mathematical notation, such that, $f(\varepsilon) = \Theta(g(\varepsilon))$ if $f(\varepsilon) = \underline{O}(\varepsilon^d)$, and $f(\varepsilon) = \overline{O}(\varepsilon^d)$. Here, \underline{O} is such that $f(\varepsilon) = \underline{O}(g(\varepsilon))$ if $|f(\varepsilon)| \geq c_2 g(\varepsilon)$ for some constant $c_2 > 0$ for all ε sufficiently small. The \overline{O} notation in an asymptotic analysis is such that $f(\varepsilon) = \overline{O}(g(\varepsilon))$ if $|f(\varepsilon)| \leq c_1 g(\varepsilon)$ for some constant $c_1 > 0$ for all ε sufficiently small.

3 MONTE CARLO SIMULATION

In CMC method, independent samples of X are generated to estimate the unknown u by an unbiased estimator for which the s and t is disconnected by (L'Ecuyer et al. 2011): $\bar{U}_{MC}^{(n)} = \frac{1}{n} \sum_{i=1}^n \psi(X_i^{(i)})$

The precision of the estimator $\bar{U}_{MC}^{(n)}$ is measured by its variance (lower value means better precision) $(S^{(n)})^2 = \bar{U}_{MC}^{(n)}(1 - \bar{U}_{MC}^{(n)})n/(n-1)$, and the confidence interval on the estimation of u is given by

$$\bar{U}_{MC}^{(n)} - c_\alpha S_{MC}^{(n)}/\sqrt{n}, \bar{U}_{MC}^{(n)} + c_\alpha S_{MC}^{(n)}/\sqrt{n} \quad (2)$$

The relative half-width $c_\alpha \frac{((S_{MC}^{(n)})^2/n)^{1/2}}{\mathbb{E}[\psi(X)]} = c_\alpha \left(\frac{1-u}{un}\right)^{1/2}$ of the confidence-interval for a confidence level α increases to ∞ when $u \rightarrow 0$ (i.e., rare-event) for a fixed n (L'Ecuyer et al. 2011, Rubino and Tuffin 2009). For a fixed RE, the number of realizations n is inversely proportional to u and the required n would increase exponentially as $u \rightarrow 0$ (Rubino and Tuffin 2009). Thus, we require more efficient techniques than CMC method for rare-event analysis.

4 IMPORTANCE SAMPLING

Importance Sampling (IS) is an advanced class of variance reduction techniques used for rare event analysis. The generic idea and the approximate zero-variance scheme are explained in the following sections.

4.1 General Description

The original probabilities \mathbb{P} of the 2^m possible configurations of X are replaced by a new probability $\tilde{\mathbb{P}}$ which gives

$$u = \mathbb{E}[\psi(X)] = \sum_{x \in \{0,1\}^m} \psi(x) \mathbb{P}[X = x] = \sum_{x \in \{0,1\}^m} \psi(x) L(x) \tilde{\mathbb{P}}[X = x]$$

where the likelihood ratio $L(x) = \mathbb{P}[X = x]/\tilde{\mathbb{P}}[X = x]$. The condition $\tilde{\mathbb{P}}[X = x] > 0$, when $\mathbb{P}[X = x] > 0$ must be met when $\psi(x) > 0$.

The unreliability is now $u = \tilde{\mathbb{E}}[\psi(X)L(X)]$ and the unbiased estimator obtained from importance sampling takes the form $\bar{U}_{IS}^{(n)} = \frac{1}{n} \sum_{j=1}^n \psi(X^{(j)})L(X^{(j)})$, where $X^{(1)}, \dots, X^{(n)}$ are s-independent copies of X distributed according to $\tilde{\mathbb{P}}$. The confidence interval over u under the new probability law $\tilde{\mathbb{P}}$ can be obtained from (2), by replacing the sample mean $\bar{U}_{MC}^{(n)}$ with $\bar{U}_{IS}^{(n)}$, and the variance $(S_{MC}^{(n)})^2$ with the sample variance $(S_{IS}^{(n)})^2$ of $\psi(X^{(j)})L(X^{(j)})$. As explained by L'Ecuyer et al. (2011), if $\tilde{\mathbb{P}}$ is the optimal probability for which the variance is reduced to zero (i.e., ideal zero-variance estimator), all the probabilities are inflated by a factor proportional to $\psi(x)$ and $\tilde{\mathbb{P}}[X = x] = \psi(x)\mathbb{P}[X = x]/u$ for all the configurations of $x \in \{0, 1\}^m$. This equation means that, for the realizations for which the system does not fail (i.e., $\mathbb{P}[X = x]$), the sampling $\tilde{\mathbb{P}}[X = x] = 0$, while for the other realizations where system fails (i.e., $u > 0$), the original \mathbb{P} is to be divided by u to obtain the optimal $\tilde{\mathbb{P}}$. However, this method is impractical because it requires the knowledge of u , the value we precisely want to compute.

4.2 Zero-Variance Importance Sampling Approximation

Under the zero-variance IS method, as described by L'Ecuyer et al. (2011), but considering the sampling of nodes instead of links, node states are sampled sequentially given the state of previously sampled nodes. Formally, if $q_i = \mathbb{P}[X_i = 0] = 1 - \mathbb{P}[X_i = 1]$ under \mathbb{P} , then q_i is changed at each step depending on the previously generated values X_1, \dots, X_{i-1} . The unreliability of the graph \mathcal{G}' , conditional on the already sampled nodes 1 to $i-1$ is given by $u_i(x_1, \dots, x_{i-1}) = \mathbb{E}[\psi(X)|X_1 = x_1, \dots, X_{i-1} = x_{i-1}]$, which also means $u_i(x_1, \dots, x_{i-1}) = q_i u_{i+1}(x_1, \dots, x_{i-1}, 0) + (1 - q_i) u_{i+1}(x_1, \dots, x_{i-1}, 1)$, and the overall unconditional unreliability of the graph can be written as $u = u_1(\emptyset)$ (L'Ecuyer et al. 2011).

If for $i = (1, \dots, m)$, q_i is replaced by

$$\tilde{q}_i \stackrel{\text{def}}{=} \tilde{\mathbb{P}}[X_i = 0 | X_1 = x_1, \dots, X_{i-1} = x_{i-1}] = q_i \frac{u_{i+1}(x_1, \dots, x_{i-1}, 0)}{u_i(x_1, \dots, x_{i-1})}, \quad (3)$$

as shown by L'Ecuyer et al. (2011), following the same arguments for the proof, this sequential IS gives a zero-variance estimator. However, it (again) requires the exact knowledge of all the functions of u_i and specifically $u_1(\emptyset) = u$, which is not practical.

Following L'Ecuyer et al. (2011), we propose to replace $u_i(\cdot)$ in (3) by an approximation $\hat{u}_i(\cdot)$, which gives

$$\tilde{q}_i = \tilde{\mathbb{P}}[X_i = 0] = \frac{q_i \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 0)}{q_i \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 0) + (1 - q_i) \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 1)}. \quad (4)$$

If $\hat{u}_{i+1}(\cdot)$ is not too far from $u_{i+1}(\cdot)$ for each i , then the variance would be reduced by a large factor. It is important to note that the network unreliability u will not change according to the order in which the nodes (or vertices) are numbered in the graph but the change of measure would depend on the ordering in the proposed algorithm. In our analysis, we found out that certain enumeration of nodes does vary the estimated unreliability \hat{u} by a very small factor and so does the value of RE. However, we do not have yet a robust heuristic to choose the ordering of the nodes which could evaluate the optimum unreliability estimate \hat{u} possible, or a correlation between the estimation or RE with the ordering.

Using this IS scheme, we can prove in an asymptotic regime where $\varepsilon \rightarrow 0$, while the graph topology is fixed, that some condition on the approximation $\hat{u}_i(\cdot)$ guarantees that BRE or even VRE are satisfied. Recall that BRE means the standard deviation of the estimator divided by the mean value σ/u is kept bounded as $\varepsilon \rightarrow 0$; in other words the sample size to get a predefined RE is independent of the rarity parameter ε . VRE means that σ/u tends to zero with $\varepsilon \rightarrow 0$: asymptotically the estimator is perfect. The conditions are the same as in L'Ecuyer et al. (2011), with nodes considered instead of links but again the arguments are exactly the same. The impact of considering nodes is more in the computation of the chosen approximation that will be described later.

- Let us suppose that for each i and $(x_1, \dots, x_i) \in \{0, 1\}^i$, $1 \leq i \leq m$, there is a constant $a_{i+1}(x_1, \dots, x_i)$ independent of ε such that

$$\hat{u}_{i+1}(x_1, \dots, x_i) = a_{i+1}(x_1, \dots, x_i) u_{i+1}(x_1, \dots, x_i) + o(u_{i+1}(x_1, \dots, x_i)). \quad (5)$$

If this condition is satisfied, then BRE holds.

- Let us define $S_1 = \{x \in \{0, 1\}^m : \psi(x) = 1 \text{ and } \tilde{\mathbb{P}}[X = x] = \Theta(1)\}$, and $S_0 = \{x \in \{0, 1\}^m : \psi(x) = 1 \text{ and } \tilde{\mathbb{P}}[X = x] = o(1)\}$.

The union $S_0 \cup S_1$ is the set of possible configurations where the system fails. The configurations in S_1 are not rare under IS, while the ones in S_0 are still rare. The required additional condition for VRE involves $x \in S_1$ only. Assuming the assumptions defined for BRE hold, and supposing $x = (x_1, \dots, x_m) \in S_1$ and for each i , one of the following three conditions is satisfied to prove the VRE property:

$$\frac{\hat{u}_{i+1}(x_1, \dots, x_{i-1}, 1)}{u_{i+1}(x_1, \dots, x_{i-1}, 1)} = \frac{\hat{u}(x_1, \dots, x_{i-1}, 0)}{u_{i+1}(x_1, \dots, x_{i-1}, 0)} + o(1), \quad (6)$$

or

$$x_i = 0, a_{i+1}(x_1, \dots, x_i) = 1, \text{ and } (1 - q_i) \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 1) = o(q_i \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 0)) \quad (7)$$

or

$$x_1 = 1, a_{i+1}(x_1, \dots, x_i) = 1, \text{ and } q_i \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 0) = o((1 - q_i) \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 1)). \quad (8)$$

5 APPLICATION AND RESULTS

In this section we illustrate the various topologies considered in our analysis along with the Data Communication System (DCS) structure. Also, we show the results obtained from the application of the approximate zero-variance IS method based on mincuts. For the purpose of our study, we modified the maxflow-mincut algorithm proposed by Ford-Fulkerson (Sedgewick and Schidlowsky 2003) for considering flow through nodes. The algorithm used thereafter is based on the one explained in Section 5.1, where the computation based on node failures is presented.

5.1 Approximation Based on Mincuts Computed From Ford-Fulkerson Maxflow-Mincut Algorithm

The proposed approximation of \hat{u}_i is to consider the probability of a *mincut with a maximal probability* where nodes (and associated links) sampled as failed are removed from the graph, and nodes sampled operational are compacted. Recall that a cut of a graph is defined as the partition of nodes of the graph into two disjoint subsets of \mathcal{G} while a mincut γ is a cut whose capacity is minimum over all the cuts of \mathcal{G} . A mincut with a maximal probability is a mincut whose probability that all nodes are failed, is computed as the product of the failure probabilities of those nodes. With such an approximation, the condition for BRE are always satisfied (similarly to L'Ecuyer et al. (2011)), and VRE can be satisfied in some cases.

The question is now, how to compute such an approximation in the case of nodes? As explained in Section 2 by (1), we parameterize the system unreliability in an asymptotic regime with respect to a rarity parameter $\varepsilon \rightarrow 0$ such that $q_i \rightarrow 0 \forall i$. Define $\varepsilon = \max_i q_i$ such that $\forall i$,

$$q_i = \varepsilon^{c_i} \quad (9)$$

with $c_i = \frac{\log q_i}{\log \varepsilon} \geq 1$. Calling c_i the *capacity* of node i , and c the capacity of the graph obtained from maxflow-mincut algorithm proposed by Ford-Fulkerson based on links, we get $c = \sum_{i \in \mathcal{C}} c_i$ for \mathcal{C} , a mincut with a maximal probability, and the corresponding probability is $q_{\mathcal{C}} = \varepsilon^c = \prod_{i \in \mathcal{C}} \varepsilon^{c_i} = \prod_{i \in \mathcal{C}} q_i$. The trick is to use the log to switch from the sum of capacities to the product of probabilities.

The Ford-Fulkerson algorithm (Sedgewick and Schidlowsky 2003) adapted for nodes differs from the case of links such that: if a node i fails (i.e., $x_i = 0$), all its associated links are useless and removed, and if a node i is considered working (i.e., $x_i = 1$) then it is removed from the graph model and its associated nodes are mutually linked to each other. This makes the algorithm for the node failure case more complex compared to link failure where a failed link can be just removed and the connecting nodes of a perfectly working link are merged. The Algorithm 1 proposed illustrates the adapted Ford-Fulkerson algorithm.

Algorithm 1 Adapted Ford-Fulkerson Maxflow-Mincut Algorithm

- 1: use *nodelist* (list) to store node names of any random graph \mathcal{G}' or \mathcal{G}
 - 2: **for** $i = 1$ to $size.nodelist()$ **do**
 - 3: assign capacity c_i to node i using (9)
 - 4: **end for**
 - 5: initialize *cap*; initialize *apath* (array) and *flow* (array) for all nodes i
 - 6: use Breadth First Search (BFS) to find a path *apath* between s and t passing *apath* and *flow* argument
 - Ensure:** BFS finds path only where *flow* can be assigned and return a value $P > 0$ if found a path
 - 7: **while** $P > 0$ **do**
 - 8: find node with minimum c_i (*mincap*) in *apath*
 - Ensure:** remaining capacity in all the nodes in *apath* is greater than *mincap*
 - 9: assign *flow* equal to the *mincap* to all nodes in *apath*
 - 10: *cap* += *mincap*
 - 11: initialize *apath* and use BFS to find another path where *flow* can be assigned and return $P > 0$
 - 12: **end while**
 - 13: **return** *cap*
-

The mincut with maximal probability problem is $\Theta(u)$ as explained by L'Ecuyer et al. (2011) and as a consequence the BRE property is satisfied in the case of node failures too. With more stricter conditions (one of the (6) or (7) or (8)) satisfied, the VRE property is also observed as illustrated by L'Ecuyer et al. (2011) for link failures. The overall procedure of the proposed zero-variance IS scheme based on mincuts is illustrated by Algorithm 2.

Algorithm 2 Approximate Zero-Variance IS using Ford-Fulkerson adapted algorithm

- 1: $L \leftarrow 1$; Starting with original graph \mathcal{G}
 - 2: **for** $i = 1$ to m until s are directly connected or completely disconnected i.e., BFS find direct path or no path
 - 3: Compress node i corresponding to $\mathcal{G}(x_1, \dots, x_{i-1}, 1)$ of Step 2 and mutually connect its neighbors in \mathcal{G}'
 - 4: find cap (mincut of maximal probability in the set \mathcal{C}_i) using Algorithm 1
 - 5: $\hat{u}_{i+1}(x_1, \dots, x_{i-1}, 1) \leftarrow \mathbb{P}[E(cap)]$;
 - 6: Erase node i corresponding to $\mathcal{G}(x_1, \dots, x_{i-1}, 1)$ of Step 2 and remove it everywhere in \mathcal{G}'
 - 7: find cap (mincut of maximal probability in the set \mathcal{C}_i) using Algorithm 1 again
 - 8: $\hat{u}_{i+1}(x_1, \dots, x_{i-1}, 0) \leftarrow \mathbb{P}[E(cap)]$;
 - 9: compute \tilde{q}_i via (4)
 - 10: generate U_i a random variate over $(0, 1)$;
 - 11: **if** $U_i < \tilde{q}_i$ **then**
 - 12: $x_i \leftarrow 0$; $L_i \leftarrow q_i / \tilde{q}_i$;
 - 13: **else**
 - 14: $x_i \leftarrow 1$; $L_i \leftarrow (1 - q_i) / (1 - \tilde{q}_i)$;
 - 15: **end if**
 - 16: $L \leftarrow L \times L_i$;
 - 17: **end for**
 - 18: **return** $Y = \psi(x_1, \dots, x_m) \times L$
-

5.2 Numerical Results

We considered four topologies for illustrating the mincut-maxprob approximation. In all the studied topologies, the nodes are sampled by order of their numbers from s to t . The nodes are homogeneous with unreliability $q_i = \varepsilon$ for $i = 1, 2, 3, \dots, m$, where $\varepsilon \in \mathbb{R}$. The metric of interest is the probability that s and t are disconnected in the following networks.

Example 1: First we consider a graph with 21 nodes and 36 links as shown in Figure 1. Computing recursively, its exact solution of the unreliability is: $u = 1 - (1 - \varepsilon)^7[(1 - \varepsilon)^3 - 3(1 - \varepsilon)^2 + 3 - 2\varepsilon]^4$, where $q_i = \varepsilon$ homogeneously assigned and u is the exact unreliability of the graph.

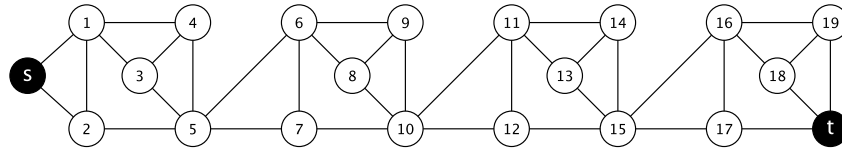


Figure 1: 21 nodes graph.

The IS scheme adapted for the case of node failures suggest that the BRE property holds while we also obtain a tight confidence interval over the estimation of u as $\varepsilon \rightarrow 0$, as shown in Table 1. The exact analytical solution is also bounded in the 95% confidence interval we obtain from the simulations, thus proving the precision of the adapted IS scheme. Also, comparing it with CMC method, for the same topology the proposed IS scheme average simulation time was 1.7962×10^{-4} seconds per iteration while for the CMC method it was 4.99×10^{-6} seconds per iteration. However, the CMC method didn't record any failure event for $\varepsilon < 10^{-6}$. Also, the work normalized variance (var_{wn}) (variance multiplied by the expected

computing time per iteration) in the IS scheme is much lower and reduces much rapidly as $\varepsilon \rightarrow 0$ for any value of ε as compared to the one obtained from MC simulations. For e.g., for a rare event $\varepsilon = 10^{-7}$, $(S_{MC}^{(n)})^2 = \hat{u}(1 - \hat{u})$ and CMC method's var_{wn} is 1.50×10^{-12} . But with zero-variance approximation based on mincuts, var_{wn} for same case is 1.80×10^{-18} which is much lower.

Table 1: Empirical results for graph with 21 nodes and $n = 10^6$.

ε	Exact Soln	Estimate	95 % CI	STD	R.E.	Time (s)
10^{-3}	3.0010×10^{-3}	3.0005×10^{-3}	$(2.9986 \times 10^{-3}, 3.0025 \times 10^{-3})$	1.01×10^{-3}	0.34	191.62
10^{-5}	3.0000×10^{-5}	2.9995×10^{-5}	$(2.9975 \times 10^{-5}, 3.0015 \times 10^{-5})$	1.00×10^{-5}	0.33	179.56
10^{-7}	3.0000×10^{-7}	2.9995×10^{-7}	$(2.9975 \times 10^{-7}, 3.0014 \times 10^{-7})$	1.00×10^{-7}	0.33	178.72
10^{-9}	3.0000×10^{-9}	2.9995×10^{-9}	$(2.9975 \times 10^{-9}, 3.0014 \times 10^{-9})$	1.00×10^{-9}	0.33	176.89

Example 2: We now take a Dodecahedron topology having 20 nodes and 30 links, as shown in Figure 2, which is often used as a benchmark for network reliability estimation techniques (Cancela et al. 2009).

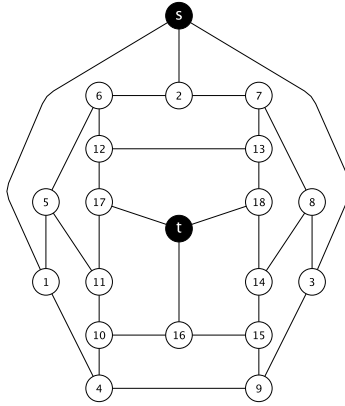


Figure 2: Dodecahedron.

We considered different ways of ordering of nodes in our analysis. In all the cases the VRE property was observed, as shown in Table 2 for one example. Comparing the results with CMC method, the average per unit computation time for the proposed IS scheme was 3.07×10^{-4} seconds while CMC method took approximately 4.70×10^{-6} seconds without recording a single failure event (for $\varepsilon < 10^{-2}$). From the estimated \hat{u} , for example for a rare event $\varepsilon = 10^{-7}$, var_{wn} obtained from CMC method is 9.40×10^{-27} and with the proposed IS scheme is much lower (1.23×10^{-59}). As $\varepsilon \rightarrow 0$, var_{wn} of the IS scheme decreases much faster than the CMC method.

Table 2: Empirical results for a single dodecahedron and $n = 10^6$.

ε	Estimate	95 % CI	STD	R.E.	Time (s)
10^{-3}	2.0061×10^{-9}	$(2.0058 \times 10^{-9}, 2.0064 \times 10^{-9})$	1.50×10^{-10}	7.45×10^{-2}	310.70
10^{-5}	2.0001×10^{-15}	$(2.0000 \times 10^{-15}, 2.0001 \times 10^{-15})$	1.31×10^{-17}	6.56×10^{-3}	315.47
10^{-7}	2.0000×10^{-21}	$(2.0000 \times 10^{-21}, 2.0000 \times 10^{-21})$	2.00×10^{-28}	1.00×10^{-7}	304.99
10^{-9}	2.0000×10^{-27}	$(2.0000 \times 10^{-27}, 2.0000 \times 10^{-27})$	2.00×10^{-36}	1.00×10^{-9}	313.17

Example 3: The third network considered is a much larger network (L'Ecuyer et al. 2011) where three dodecahedrons of Figure 2 are juxtaposed in a parallel configuration as shown in Figure 3. The source s and terminal t of Figure 2 are merged and represented by a single node, s and t respectively. The topology has 56 nodes and 90 links.

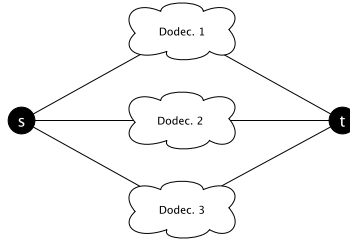


Figure 3: Three dodecahedrons connected in parallel.

We consider a homogeneous case such that all nodes have the same unreliability ε and the goal is to compute the unreliability that the s and t are not connected. The unreliability obtained here is a cube of the unreliability of a single dodecahedron for the case of node failures (L’Ecuyer et al. 2011). It is observable that the unreliability estimates in the case of node failures, as shown in Table 3 are of the same order of magnitude for both Example 2 (single dodecahedron) and Example 3 (three parallel dodecahedrons) as it is for the case of link failures obtained by L’Ecuyer et al. (2011). The empirical results from Table 3 show that the VRE property holds. Comparing with CMC method, for example for $\varepsilon = 10^{-7}$ rare event, var_{wn} for CMC method is 1.18×10^{-67} (average per run computation time = 1.47×10^{-5} seconds). For the same case, with the IS scheme var_{wn} is 3.95×10^{-133} (average per run computation time = 6.16×10^{-3} seconds). The var_{wn} for IS scheme decreases (as $\varepsilon \rightarrow 0$) much rapidly compared to the one that is obtained from CMC method.

Table 3: Empirical results for three dodecahedrons connected in parallel and $n = 10^6$

ε	Estimate	95 % CI	STD	R.E.	Time (s)
10^{-3}	8.0739×10^{-27}	$(8.0714 \times 10^{-27}, 8.0764 \times 10^{-27})$	1.27×10^{-27}	1.57×10^{-1}	6679.12
10^{-5}	8.0005×10^{-45}	$(8.0003 \times 10^{-45}, 8.0007 \times 10^{-45})$	1.03×10^{-46}	$1/28 \times 10^{-2}$	6738.80
10^{-7}	8.0000×10^{-63}	$(8.0000 \times 10^{-63}, 8.0000 \times 10^{-63})$	8.00×10^{-66}	1.00×10^{-3}	6249.11
10^{-9}	8.0000×10^{-81}	$(8.0000 \times 10^{-81}, 8.0000 \times 10^{-81})$	1.39×10^{-89}	1.73×10^{-9}	5704.38

5.3 Practical Case Study: Data Communication System

We now consider a real Data Communication System (DCS), a part of a large scale passenger rail system *Communication Based Train Control* (CBTC). The role of a DCS is to carry without hindrance the data between various other rail systems ensuring end-to-end communication. It consists of reliable and redundant communication paths, as shown by RED and BLUE in Figure 4. In Figure 4, the links represent wired or wireless communication channels between ground-to-ground or ground-to-rolling stock (train), respectively. The nodes represent ethernet or electrical switches, routers, servers, radio equipment, modems, etc. The red and blue components (links and nodes) are in pairs, and are in UP state all the time such that there is no switching of functioning if one of them fails. This also adds a complexity of undetected failure. However, the redundancy of red and blue makes certain that there are two independent communication paths available all the time. In Figure 4, the train is considered as the source s (outermost node) and the Zone Controller server is considered as the terminal t (innermost node). The circular topology is for the purpose of simplicity for the readers. In the DCS, the outer circle of nodes represent the trackside equipment which communicates directly with the train through overlapping wireless radio access points’ coverage. A failure of more than three consecutive pairs of red and blue nodes will make the s and t disconnected.

In our analysis, for simplicity, there is only one train and it is considered to be at a fixed position as shown in Figure 4. Practically there are more than one train present (i.e., multiple sources s) along the

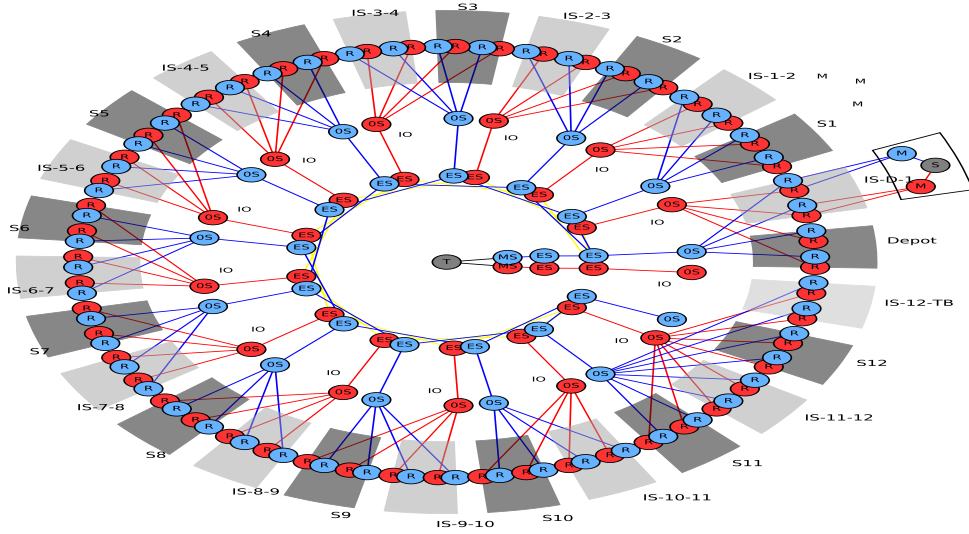


Figure 4: Data Communication System (DCS).

outer circle. Also, we consider the network as a homogeneous network where all the nodes have the same unreliability ε . There are 164 nodes and 169 links in the graph model.

The empirical results from the Table 4, show that the BRE property holds when $\varepsilon \rightarrow 0$, and we obtain tight bounds over the estimated values of the unreliability u . The results from this static model give the steady-state availability of the system. From the estimated \hat{u} , for example when $\varepsilon = 10^{-7}$, the var_{wn} for CMC method is 6.17×10^{-18} (average per run computation time being 3.87×10^{-5} seconds). For the proposed IS scheme, the var_{wn} is 8.92×10^{-28} (average per run computation time being 7.82×10^{-3}), which is much lower and also decreases much rapidly as $\varepsilon \rightarrow 0$.

Table 4: Empirical results for DCS topology.

ε	Estimate	95 % CI	STD	R.E.	Time (s)
10^{-3}	1.5954×10^{-5}	$(1.5888 \times 10^{-5}, 1.6020 \times 10^{-5})$	3.38×10^{-5}	2.12	7485.66
10^{-5}	1.5968×10^{-9}	$(1.5902 \times 10^{-9}, 1.6034 \times 10^{-9})$	3.38×10^{-9}	2.11	7426.27
10^{-7}	1.5968×10^{-13}	$(1.5902 \times 10^{-13}, 1.6034 \times 10^{-13})$	3.38×10^{-13}	2.11	9168.73
10^{-9}	1.5968×10^{-17}	$(1.5902 \times 10^{-17}, 1.6034 \times 10^{-17})$	3.38×10^{-17}	2.11	7802.55

The above analysis is done on a MacBook with a 1.8 GHz Intel Core i5 processor with Java implementation. The zero-variance IS scheme is more computationally burdensome compared to CMC methods, as it needs to find two mincuts with maximal probability at each step of the sampling process (L'Ecuyer et al. 2011) using a Ford-Fulkerson adapted algorithm. However, the method estimates the unreliability u with a higher accuracy (variance reduction) at the expense of increased computation time. This is a trade off between choosing a more precise estimate or a faster estimate with huge variance. With respect to rare event analysis, if we quantify the gain in terms of work normalized variance var_{wn} , which gives the estimate of variance reduction with respect to the cost (i.e., time), the proposed zero-variance IS scheme is highly efficient compared to the CMC method.

6 CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

We describe in the paper an adapted zero-variance approximate IS methodology for considering node failures for the purpose of static network reliability estimation. The sequential sampling of nodes, as done by L'Ecuyer et al. (2011), reduces the variance by a large factor. We prove that the methodology explained here and by L'Ecuyer et al. (2011), works for the case of node failures also and illustrate its efficiency on a real network, a Data Communication System used in urban train control.

As a possible extension of the method, we will try to improve the possibility of finding a robust heuristic for ordering of the nodes which gives the best possible estimates of unreliability with lowest possible values of RE. Another possibility is to reduce the computational time using graph reduction techniques. Another possibility is to extend this methodology for application on repairable dynamic systems where each component could be in more than one state (for e.g., working, failed, detected/undetected failure, not working at full capacity, etc.), including logistics, that could prove its usefulness in a more realistic manner for problems such as simulation optimization, run time of experiments, accuracy of estimates etc., as posed by Dersin and Valenzuela (2012).

REFERENCES

- Ball, M. O. 1986. "Computational Complexity of Network Reliability Analysis: An Overview". *IEEE Transactions on Reliability* 35 (3): 230–239.
- Cancela, H., M. E. 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, Chapter 7, 145–170. Wiley Online Library.
- Dersin, P., and R. Valenzuela. 2012. "Application of non-Markovian stochastic Petri Nets to the modeling of rail system maintenance and availability". In *Simulation Conference (WSC), Proceedings of the 2012 Winter*, 1–12. IEEE.
- Gertsbakh, I., Y. Shpungin, and R. Vaisman. 2014. "Network reliability Monte Carlo with nodes subject to failure". *International Journal of Performability Engineering* 10 (2): 163–172.
- L'Ecuyer, P., J. Blanchet, B. Tuffin, and P. Glynn. 2010. "Asymptotic robustness of estimators in rare-event simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 20 (1): 6.
- L'Ecuyer, P., M. Mandjes, and B. Tuffin. 2009. "Importance sampling in rare event simulation". In *Rare event simulation using Monte Carlo methods*, edited by G. Rubino and B. Tuffin, Chapter 2, 17–38. Wiley Online Library.
- L'Ecuyer, P., G. Rubino, S. Saggadi, and B. Tuffin. 2011. "Approximate zero-variance importance sampling for static network reliability estimation". *Reliability, IEEE Transactions on* 60 (3): 590–604.
- Rubino, G. 1998. "Network reliability evaluation". *State-of-the art in performance modeling and simulation*:275–302.
- Rubino, G., and B. Tuffin. 2009. "Introduction to rare event simulation". In *Rare event simulation using Monte Carlo methods*, edited by G. Rubino and B. Tuffin, Chapter 1, 1–13. Wiley Online Library.
- Sedgewick, R., and M. Schidlowsky. 2003. "Algorithms in Java, Part 5". *Addison-Wesley, Boston*.

AUTHOR BIOGRAPHIES

AJIT RAI is a Ph.D. candidate at the University of Rennes 1, France under an industrial agreement for training through research (CIFRE) between ALSTOM and INRIA (the French National Institute for Research in Computer Science and Control) at Rennes, France. He holds a joint european Masters degree in nuclear engineering from Polytechnic University of Catalonia, Spain and Grenoble Institute of Technology, France, and a Bachelors degree in Mechanical Engineering from Gautam Buddh Technical University, India. His research interests include importance sampling methods for rare-event analysis and their development to

evaluate system availability including logistical aspects. His email address is ajit.rai@transport.alstom.com.

RENE C. VALENZUELA received his Ph.D. degree in Aerospace Engineering from the Georgia Institute of Technology (2013). He is currently the Systems Methods & Tools RAM Leader within ALSTOM Transport Information Solutions RAM Centre of Excellence. His research interests include developing and applying software reliability models and application of simulation models to estimation of reliability and quality of service metrics. He has presented his work in ESREDA, ESREL and WSC conferences. His email address is rene.valenzuela@transport.alstom.com.

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. His research interests include developing Monte Carlo and quasi-Monte Carlo simulation techniques for the performance evaluation of telecommunication systems and telecommunication-related economical models. He has published more than one hundred papers on those issues. He is currently Area Editor for *INFORMS Journal on Computing* and Associate Editor for *ACM Transactions on Modeling and Computer Simulation*. He has written or co-written three books (two devoted to simulation): *Rare event simulation using Monte Carlo methods* published by John Wiley & Sons in 2009, *La simulation de Monte Carlo* (in French), published by Hermes Editions in 2010, and *Telecommunication Network Economics: From Theory to Applications*, published by Cambridge University Press in 2014. His email address is bruno.tuffin@inria.fr.

GERARDO RUBINO is a senior researcher at INRIA in Rennes, France, where he is the leader of the DIONYSOS (Dependability, Interoperability and perFormaNce analySiS of netwOrkS) team. His research interests are in the quantitative analysis of computer and communication systems, mainly using probabilistic models. He also works on the quantitative evaluation of perceptual quality of multimedia communications over the Internet. He co-edited the book *Rare event simulation using Monte Carlo methods* published by John Wiley & Sons in 2009. He is a member of the IFIP WG 7.3. His email address is gerardo.rubino@inria.fr.

PIERRE DERSIN received his Ph.D. degree from the Massachusetts Institute of Technology (MIT) in 1980 in Electrical Engineering. He joined ALSTOM Transport in 1990 and since then he has occupied several positions, mainly involved with RAMS and Maintenance, including R&D Manager of the Service business, and is now RAM Director in Alstom Transports Information Solutions (i.e. railway signaling and communication) product line. He has contributed a number of communications and publications in conferences and journals (including IEEE, ESREL 92, the 1995 RAMS Symposium and French Lambda-Mu conferences), in the fields of RAMS, automatic control and electric power systems. He is a member of the IEEE Reliability Society AdCom and the IEEE Future Directions committee. His email address is pierre.dersin@transport.alstom.com.