Lessons-learned of using Monte Carlo method with Important Sampling In Fault Tree Quantification

Heejong Yoo^a, Gyunyoung Heo^b

^a Kyung Hee University, Yongin-si, Republic of Korea, kreacher@khu.ac.kr ^b Kyung Hee University, Yongin-si, Republic of Korea, gheo@khu.ac.kr

Abstract: In the quantitative evaluation of fault trees (FTs) and event trees (ETs) during a level 1 PSA, the combinatoric process of ETs and FTs are necessary to deal with Boolean logic. Most of quantification processes use minimal cut sets, which needs additional process of gaining the minimal cut sets and its validation. Therefore, the need of developing methods that is free of minimal cut sets may have benefits from these viewpoints. While there were some attempts to gain the top vent probability of FTs by using the Monte Carlo method, which could be free from using minimal cut sets by setting a different algorithm, the time and computational cost for using the Monte Carlo method is always the main issue. In order to reduce computational resource and having its strong point in variance reduction, the most frequently used application for the Monte Carlo method is importance sampling. This paper suggests an algorithm of implying importance sampling, a general method used to reduce the cost for the Monte Carlo method, in order to quantify FTs, and show both the application and limitations of importance sampling. An example FT is given in the paper to show the application process.

1. INTRODUCTION

Probabilistic safety assessment (PSA) projects are widely being in progress worldwide, while level 1 PSA projects are the mainstream due to sequential progressing [1]. Level 1 PSA has four steps: initial event declaration, accident scenario analysis, accident scenario quantification, and result analysis [2]. In the quantification process, the sequences in ETs are calculated easily by the product of the probability of each system to succeed or fail. However, FTs, which has the probability of each system as the result of quantification, needs additional techniques. This is due to the logic system used in FTs.

Methods to gain the exact result for FT quantification is by using the truth table or the binary decision diagram analysis [3]. These methods consider every possible combination of basic events inside a FT, which requires numerous calculations. In FTs used in nuclear engineering, there are roughly 3,000 basic events for each, leading to 2³⁰⁰⁰ calculations. Due to the computational cost needed, these calculations are not held on in real applications. While there are adjustments available in order to reduce these computational costs, these lead to decrease in accuracy. The most frequently used methods in FT quantification the methods using minimal cut sets. Minimal cut sets are the minimal combination of component failures that causes system failure. These methods have its strong point at fast calculation, mostly in a few seconds for examples of rare event approximation (REA) and minimal cut upper bound (MCUB).

Although there is a strong point for the minimal cut set methods, there are some considerations to be clarified [4]. In order to use minimal cut sets, gaining the minimal cut sets from the FTs are needed. As FTs used in nuclear engineering is becoming complex in the logic and becoming larger in size, gaining the minimal cut sets are requiring more time and effort, especially the recent PSA research are focusing on multi-units. Minimal cut set methods use adjustments and assumptions for fast calculation, and this leads to increased errors for the calculations containing events with high probability. Also, conventional methods all uses minimal cut sets, leading to the need of using methods without using minimal cut sets for validation.

Monte Carlo methods are also introduced in FT quantification [5,6]. With the known work needed to reduce the time cost, there are a number of strong points Monte Carlo method have over the conventional methods using minimal cut sets. Although Monte Carlo method also favors minimal cut sets in the quantification, the calculation is available even without deriving minimal cut sets, reducing the effort to gain minimal cut sets. Using Monte Carlo method is also not affected by events having high probabilities, which causes high-level of errors in the conventional methods. Quantification using Monte Carlo method without using minimal cut sets would be useful in the cross-validation process. In order to maintain these strong points, this paper suggests a branch method of using Monte Carlo method for FT quantification. Particularly applying importance sampling in the Monte Carlo method was investigated to reduce the resource cost and variance. With the algorithm used, the pros and cons for using Monte Carlo method with importance sampling will be elaborated in the latter part of the paper, and examples from the AND and OR gates are shown.

2. METHODOLOGY

2.1. Boolean Logic in FT

FTs use Boolean logic in order to express the correlations between events and gates. Events are the basic components, having its unique probability to fail, and gates are used to group lower level events or gates. How the grouping is done is where Boolean logic is used, and are expressed in the gates with symbols. Examples of gates are AND gate, OR gate, NOT gate, NAND/NOR gates, XAND/XOR gates, and k-out-of-n gates. Figure 1 shows the gates most frequently used in FTs.



Figure 1: Frequently-used Boolean Logic Gates

Among the gates, the most frequently used ones are AND and OR gates. Calculations of AND and OR gates are similar to the calculation of sets, where AND gates follow the calculation of intersections and OR gates follow the calculation of unions. AND gates are calculated by the product of each probability of the lower level events or gates. The calculation of OR gates are first the sum of each probability of the lower levels and screening out the over-calculated intersections, or by adjustments, just simply the summation of each probability. The calculation of AND and OR gates are expressed in formulas (1) and (2), where TOP, A, and B are the probabilities of each gate (TOP) and events (A, B):

$$TOP = A \times B \tag{1}$$

$$TOP = A + B - AB \approx A + B \tag{2}$$

2.2. Exact Solution Calculation

Methods to gain the exact solution for the results of FT quantification are truth table or binary decision diagram (BDD) method. The calculation process of the two methods are quite similar, and the method of using the truth table will be described in this paper.

There are two states, success or failure, for each basic event possible. Considering every single basic events and the success or failure for each, being able to show every possible combination of basic events to succeed or fail is possible. The total number of combinations would be the power of two, and it is possible to examine whether each combination would result in the failure of the top event. For the combinations that causes top event failure, the probabilities for each basic events would use the probability as it is, while the succeeded basic events would use the value of the probability subtracted from one. Since the combinations are independent, the probabilities of each combination are summed up to gain the exact solution. Figure 2 and Table 1 shows an example of a simple FT quantified with truth table.





 Table 1: Truth Table for Example FT

Combination No.	Combination Def	Probability of TOP	System Operation
	(System States)		
1	A(S)B(S)C(S)D(S)	0.5832	S
2	A(S)B(S)C(S)D(F)	0.1458	S
3	A(S)B(S)C(F)D(S)	0.0648	S
4	A(S)B(S)C(F)D(F)	0.0162	S
5	A(S)B(F)C(S)D(S)	0.0648	S
6	A(S)B(F)C(S)D(F)	0.0162	S
7	A(S)B(F)C(F)D(S)	0.0072	S
8	A(S)B(F)C(F)D(F)	0.0018	S
9	A(F)B(S)C(S)D(S)	0.0648	S
10	A(F)B(S)C(S)D(F)	0.0162	S
11	A(F)B(S)C(F)D(S)	0.0072	S
12	A(F)B(S)C(F)D(F)	0.0018	S
13	A(F)B(F)C(S)D(S)	0.0072	S
14	A(F)B(F)C(S)D(F)	0.0018	F
15	A(F)B(F)C(F)D(S)	0.0008	F
16	A(F)B(F)C(F)D(F)	0.0002	F
Total		1	

The number of combinations to be considered is the power of two depending on the number of basic events, and additional calculation is needed for the combinations having the system operation of failure. FTs used in a power plant modeling contains over 3,000 basic events, which leads to the number of combinations to be considered to be 2 to the 3,000th power, with additional calculations to gain the probability of the combinations that fail. The computational cost needed for this work makes the truth table unable to be applied in the real use in FT for nuclear applications.

2.3 Minimal Cut Set Methods

Cut sets are the combinations of events that would eventually cause the system to fail, such as the combination of having every single event to fail. Minimal cut sets are the combinations of events in the minimum that would eventually cause the system to fail. In other words, combinations of events that are included in the minimal cut sets to fail will lead to the failure of the system regardless of other events' success or failure. With the example FT in Figure 2, the cut sets would be ABC, ABD, and ABCD, while the minimal cut sets would be ABC and ABD. When the combinations of ABC or ABD fails, the success or failure of D or C, respectively, would not be in consideration for the failure of the system, which makes the two combinations minimal cut sets. Formula (5) shows the expression of minimal cut sets in the example FT in Figure 2, where C_1 and C_2 are the minimal cut sets, and A, B, C and D represents each event to fail:

$$C_1 = (A, B, C), \ C_2 = (A, B, D)$$
 (5)

FT quantification using minimal cut sets, or minimal cut set method, is a frequently used method in the calculation process. The strong points of minimal cut set methods are in the simplification of the calculation and in the failure analysis of the system. Gaining the minimal cut sets makes it able to organize the scenario, which generally leads to learning the potential weaknesses of the system. Using the minimal cut sets, the probability of a top event would be the union of each probability of each minimal cut set, while the probability of each minimal cut sets is simply the product of each event to success or fail. The calculation of union for each probability of each minimal cut sets basically follows the Boolean calculation rule, which is also called the inclusion-exclusion principle, though adjustments of omitting lower terms are made for faster calculation. Rare event approximation (REA) uses only the first term, which would simply sum up the probability for the system to succeed subtracted from one. Results from MCUB are similar to the calculation until the third term of the inclusion-exclusion principle. The calculation of minimal cut sets without any adjustments for the example FT of Figure 2 is shown in formula (6) with the laws of Boolean algebra, while formulas (7) and (8) shows the calculation using REA and MCUB, respectively:

$$TOP = ABC + ABD - ABC \times ABD = ABC + ABD - ABCD$$
(6)
$$TOP = ABC + ABD$$
(7)

$$TOP = 1 - (1 - \overline{A}\overline{B}\overline{C}) \times (1 - \overline{A}\overline{B}\overline{D})$$
(8)

2.3. Monte Carlo Method

Monte Carlo method is an easy and useful tool for engineering problems. This method is proposed for the purpose of solving engineering problems that are unable to calculate using conventional mathematics. Numerical analysis is performed, using random numbers to gain the result of the equation, and adjust the counted results to gain an approximate solution to problems. Monte Carlo method is also used for cross-checking with other calculation mechanics even where the method is not strictly needed, mainly for validation issues.

Accuracy of the result gained by Monte Carlo method depends on the number of iterations held. Increased number of iterations leads to variance reduction, which leads to increased accuracy. Monte Carlo method could be expressed as formula (9):

$$\hat{\rho} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{Y>a} \left(Y^i \right), \qquad Y^i \sim p(y) \tag{9}$$

where $\hat{\rho}$ is the probability by using Monte Carlo method, p(y) is the distribution of the random number generation, N is the number of iterations, Y^i is the *i*th random number, a is the in comparison with the random number, $1_{Y>a}(Y^i)$ is the function indicating that when the random number meets the condition in question, the output is zero, and the output is one when the random number does not fit in the condition.

The variance, or error, of the Monte Carlo method is given by the correlation of the probability gained by the Monte Carlo method and the number of iterations used for the Monte Carlo method. Generally, a Monte Carlo method uses a binomial distribution, where the results for the function $1_{Y>a}(Y^i)$ is zero or one, and the results follows a normal distribution. When approximating a binomial distribution with a normal distribution, the percent error for the Monte Carlo method is given by formula (10) developed by Shooman [7]:

$$\% \ error = 200 \sqrt{\frac{1-p}{np}} \tag{10}$$

Where p is the estimated probability of the Monte Carlo method and n is the sample size, or the number of iterations used for the Monte Carlo method.

2.4. Importance Sampling

Although Monte Carlo method has its strong points in solving engineering problems that are unable by conventional mathematics, the number of iterations required for an accurate result may cause overrated time consumption. Since Monte Carlo method gains the count of a certain value that satisfies the given criteria, to gain a reasonable result for the numerical analysis typically requires an iteration that could gain a hundred times of counts. For instance, if there is a known target value, 0.0001 (10^{-4}) , and the Monte Carlo method gains one count out of 10,000 (10^{4}) iterations, the required number of iterations for an acceptable result would be a million (10^{6}) times. Moreover, solutions for engineering problems are usually unknown, which would lead to the need of more iterations, since the required iteration number is unknown.

In order to reduce the number of iterations needed, Monte Carlo method is used alongside with other techniques, while the most frequently used one is importance sampling. Importance sampling is a technique that reduces the interval of random number generation, gain an increased count due to the reduced random number generation interval, and adjust the final result by multiplying a weighting factor based on the reduced random number interval. By applying importance sampling, gaining increased number of counts with reduced iteration numbers is possible, and leads to reduced time. Importance sampling also has its strong point in accuracy, since the increased number of counts has the same effect as increased iterations, which means reduced variation of the results and higher accuracy. Monte Carlo method with importance sampling is expressed as formula (11):

$$\hat{\rho}^{IS} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{Y > a} \left(Y^i \right) \frac{p(\tilde{Y}^i)}{q(\tilde{Y}^i)}, \qquad \tilde{Y}^i \sim q(y)$$
(11)

where $\hat{\rho}^{IS}$ is the probability by using Monte Carlo method, p(y) is the distribution of the random number generation, q(y) is the distribution of the random number generation for the importance sampling, N is the number of iterations, \tilde{Y}^i is the *i*th random number, a is the in comparison with the random number, $1_{Y>a}(\tilde{Y}^i)$ is the function indicating that when the random number meets the condition in question, the output is zero, and the output is one when the random number does not fit in the condition.

3. CASE STUDY

3.1. FT quantification using Monte Carlo Method

3.1.1. Monte Carlo Method in FT Quantification

In FT quantification, Monte Carlo method uses the probability of each event to determine the success or failure of each event. The distribution of the random number generation is uniform distribution. If the random number is larger than the given probability, the event is determined to be succeeded, and the event is treated to have failed when the random number is smaller than or equal to the given probability. The success or failure of the lower events or gates is calculated to the upper gates based on Boolean logic, which leads to the final result to be the success or failure of the top event. This process is iterated, and failure of the top event for a single iteration is counted. The final probability for the system to fail is calculated by the number of counts over the total number of iterations performed.

3.1.2. FT Quantification methods comparison

A simple AND and OR gate FT model was set in order to show the difference between minimal cut set methods and Monte Carlo method. Each event, A and B, were set to have 0.1 each and 0.6 each to determine the difference of the result when using the truth table, Monte Carlo method, REA, and MUCB, each. Figure 3 shows the FT model sets for basic AND and OR gates, and Figure 4 shows the comparison of the results with the variance reduction as the iteration is proceeded of all cases given as an individual graph in Figure 5.



Figure 3: Basic AND and OR Gates with Probabilities for Case Study



Figure 5: Variance Reduction to Number of Iteration for Cases 1 and 2

Case 1 covers the cases for AND and OR gates where the event probabilities are classified as rare events, having 0.01 each. For Cases 1-1 and 2-1, which are the OR gates, the results for Monte Carlo method is shown to have an accurate result compared to the REA method. Especially, in Case 2-1, the probability calculated by REA method has a result over one, which is out of the question. The cases for AND gates, Cases 1-2 and 2-2, show that a reasonable calculation result could be gained by Monte Carlo method when a sufficient number of iterations is set. In all cases, MCUB method showed to have a precise result, and this is due to the fact that the inclusion-exclusion principle applied in the cases are a second-order term, and there are no delete-terms for the calculation of MCUB.

It is shown that in calculations for AND gates, which is the product of each probability, the conventional methods using minimal cut sets are accurate, while the calculation of OR gates are affected by the terms omitted in the inclusion-exclusion principle, having a known error. The variance, or error, of each case was given in log scale and in error percent, showing a continuous decrease as the number of iterations increased. Monte Carlo method shows its potential in a more accurate calculation in OR gate calculation when a sufficient number of iterations is possible.

3.2. FT quantification with applying Importance Sampling

3.2.1. Importance Sampling in FT Quantification

In order to reduce the time consumption when using Monte Carlo method, importance sampling could be applied. In the Monte Carlo method, the random number generation was done to determine the success or fail of the events. The probability for each event is usually classified as rare events, having probabilities under 0.01. For an example of an event having a probability of 0.01When the random number generation iteration is reduced from 0 to 1 to 0 to 0.1, the number of counts would increase ten times, which would give advantage in the number of iterations needed and in reduced variance in the result. This could also be applied for events having relatively high probability, while the reduced interval for the generator would be only for a small portion.

From the case studies from Section 3.1, it is shown that the AND gate calculations are relatively simple than OR gate calculations. To determine the pros and cons in applying importance sampling for both gates, the following sections covers importance sampling applied for each gates separately, and examine a FT combined with AND and OR gates. Figure 6 shows the simple AND and OR gate FTs that is used in the following sections.

Figure 6: AND and OR Gates used for Importance Sampling Analysis



3.2.2. AND Gates with Importance Sampling

Cases 3-1 and 4-1 in Figure 6 shows AND gates with different probability values. The different values are used to determine the weighting factor used for adjustments for the importance sampling results. Monte Carlo method with and without importance sampling were performed, with the initial generator interval from 0 to 1 and the adjust random number generation from 0 to 0.1. Figure 7 shows the result of the quantification for cases 3-1 and 4-1, where the importance sampling result is applied with the weighting factor of 0.01, and Figure 8 shows the variance reduction as the number of iterations increases.



Figure 7: AND Gate Quantification Results with and without Importance Sampling

Event Prob. 0.2 & 0.2, AND Gate

Event Prob. 0.2 & 0.3, AND Gate

Figure 8: Variance Reduction for AND Gate Quantification with/without Importance Sampling



The weighting factor for importance sampling was set as 0.01 based on the fact that the reduced interval for each event, 0.1, were multiplied, leading to the weighting factor to the square of 0.1, 0.01. This is shown to be true based on the result of the quantification, and the result also shows that applying importance sampling requires less number of iterations and the variance of the result is lower, being one-tenth with importance sampling. In conclusion, AND gates in sequential order is proved to be calculated by the product of the weighting factors in the lower level. 3.2.3. OR Gates with Importance Sampling

Cases 3-2 and 4-2 in Figure 9 shows OR gates with different probability values. Similar to AND gates, the different values are for the determination of the weighting factors for importance sampling. The random number generator is also 0 to 0.1 and 0 to 1 for Monte Carlo method with and without importance sampling, respectively. Figure 9 is the quantification results for cases 3-2 and 4-2, where the weighting factor is applied separately for each, while Figure 10 compares the variance reduction by applying importance sampling.





Figure 10: Variance Reduction for OR Gate Quantification with/without Importance Sampling



The weighting factor for cases 3-2 and 4-2 is 0.11 and 0.11227, respectively. The values are derived from the probabilities of the top event with and without importance sampling. For case 3-2, the probability for the initial top event, which is the OR gate calculation for two 0.02 events, and the probability for the importance sampling top event, which is the OR gate calculation for two 0.2 events, are directly applied to the weighting factor. It is shown that there were no significant patterns in the derivation of weighting factors for OR gates when simple OR gates quantification is performed.

3.3. Result Analysis

Monte Carlo method showed that the method is not affected by events with high probabilities. Cases 1 and 2 showed that in low probabilities, the Monte Carlo method and the minimal cut set methods show a reasonable accuracy compared to the accurate answer, the value gained by the truth table. However, the accuracy greatly differed when the event probabilities were relatively high, not being rare events. For the calculation of OR gates for events with high probabilities, which is in case 2-1, the calculation using minimal cut set methods showed to have an inaccurate solution, having a probability of over one, compared to Monte Carlo method, still having a relatively similar probability.

In the use of importance sampling, the technique showed to have a strong point in the reduced amount of iterations needed for a reasonable solution with a reduced variance, shown in cases 3 and 4. Adjustments were made for the use of importance sampling due to the increased amount of counts. In case 3, which is the use of AND gates, the weighting factor of the upper gate is determined to be the product of the weighting factor of the lower gates or events. On the contrast, case 4, which is the use of OR gates, did not show a specific pattern in the weighting factor, and the only way to gain the weighting factor is by comparing the actual values of the upper gate probability with and without importance sampling. For a specific insight for the weighting factors, Table 2 is given, showing the counts and results of Monte Carlo simulation with and without importance sampling for cases 3 and 4.

Case	With/without Importance Sampling	Gate	Lower Event Probability	Iteration	Counts	Gate Probability Regarding Counts
3-1	Monte Carlo simulation	AND	0.02 & 0.02	1E+9	400,312	4.003E-4
3-1	Importance sampling	AND	0.02 & 0.02	1E+9	40,000,119	4.000E-2
3-2	Monte Carlo simulation	OR	0.02 & 0.02	1E+9	39,600,050	3.960E-2
3-2	Importance sampling	OR	0.02 & 0.02	1E+9	360,005,519	3.600E-1
4-1	Monte Carlo simulation	AND	0.02 & 0.03	1E+9	599,866	6.000E-4
4-1	Importance sampling	AND	0.02 & 0.03	1E+9	59,998,449	6.000E-2
4-2	Monte Carlo simulation	OR	0.02 & 0.03	1E+9	49,940,690	4.994E-2
4-2	Importance sampling	OR	0.02 & 0.03	1E+9	439,989,333	4.400E-1

 Table 2: Counts and Results Obtained from Monte Carlo Simulation

By comparing the gate probabilities regarding the counts, it is able to figure out the weighting factor. The reduced interval for importance sampling for every case is one-tenth for each lower event probability. For cases 3-1 and 4-1, the AND gate showed that the counts increased by a hundred times, and the gate probability with the counts also increased by that rate. Other results for AND gates that are not shown in this paper also showed that the reduced amount of interval for importance sampling were multiplied to gain the weighting factor, or the reduction needed for the increased count, of the upper gate probability.

The calculation of OR gates with and without importance sampling, however, showed a different pattern. The counts increased in a pattern that could not be explained by just simply adding the two lower event probabilities, and were slightly lower than the added value. This result was due to the exclusive term of the calculation process. Without importance sampling, the exclusive terms in cases 3-2 and 4-2 are the product of the lower event probabilities, which are 4E-4 and 6E-4 respectively, while the exclusive terms with importance sampling are the product of the lower event probabilities with the reduced interval, which becomes 4E-2 and 6E-2 respectively. This leads to the conclusion that for the calculation of OR gates with the use of importance sampling, the probability of the upper event with and without the application of the reduced interval are required to gain the exact weighting factor.

4. CONCLUSION

Along with conventional methods of FT quantification using minimal cut sets, the usage of Monte Carlo method has its potential for the quantification process, shown by the case studies until now. Monte Carlo method does not use minimal cut sets and is not affected by the initial probabilities of events, having its strong points when there are a reasonable number of iterations. However, the number of iterations needed for a reasonable result is unknown and is relatively large, requiring techniques to reduce the time consumption.

While importance sampling technique is frequently used for the usage of Monte Carlo method, applying importance sampling for FT quantification proved out to have difficulties. In the quantification for AND gates, there were no problems, since the calculation of the weighting factor is simply the product of the reverse of each reduced amount of the interval. However, in the calculation of OR gates, the weighting factor is shown to not have a simple rule, and the weighting factor is proportional to the exact results of quantification for both with and without importance sampling. Other techniques for the time reduction would be possible and should be preceded for future studies.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government (MSIP: Ministry of Science, ICT and Future Planning) (No. NRF-2021M2D2A1A02044210) and the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission(NSSC) of the Republic of Korea. (No. 2103081)

References

[1] K. Silva, S. Sugawara, "*Reconsideration of PRA Framework - Addressing Level 3 PRA Coverage and Multi-unit Issues*," PSAM 14 (2018).

[2] C. Park and J. Ha, "Probabilistic Safety Assessment," Brain Korea, 2008, Korea.

[3] M. Modarres, M. Kaminskiy, and V.Krivtsov, "*Reliability Engineering and Risk Analysis: A Practical Guide*", CRC Press, 2016, USA.

[4] S. H. Han, "A top-down iteration algorithm for Monte Carlo method for probability estimation of a fault tree with circular logic", Nuclear Engineering and Technology, 50, pp.854-859 (2020).

[5] K. D. Rao, V. Gopika, V. V. S. S. Rao, H. S. Kushwaha, A. K. Verma and A. Srividya, "*Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment*," Reliability Engineering & System Safety, 94(4), pp.872-883 (2009).

[6] E. Gascard and Z. Simeu-Abazi, "Quantitative Analysis of Dynamic Fault Trees by means of Monte Carlo Simulations: Event-Driven Simulation Approach," Reliability Engineering & System Safety, 180, pp.487-504 (2018).

[7] A. H. Ang, W. H. Tang, "Probability Concepts in Engineering Planning and Design Volume II Decision, Risk, and Reliability," John Wiley & Sons Inc, 1984, USA.