A Novel Risk Assessment Method Based on Hybrid Algorithm for SCADA

Chen Yenan[1], Lu Tinghui[2], Li Linsen[1(✉)], and Zhang Han[1]

[1] School of Cyber Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
{chenyenan10,lsli}@sjtu.edu.cn
[2] Guangdong Power Grid Co., Ltd. Jiangmen Power Supply Bureau, Jiangmen, China

Abstract. With the frequent occurrence of cyber attacks in recent years, cyber attacks have become a major factor affecting the security and reliability of power SCADA. We urgently need an effective SCADA risk assessment algorithm to quantify the value at risk. However, traditional algorithms have the shortcomings of excessive parsing variables and inefficient sampling. Existing improved algorithms are far from the optimal distribution of the sampling density function. In this paper, we propose an optimal sampling algorithm and a selective parsing algorithm and combine them into an improved hybrid algorithm to solve the problems. The experimental results show that the improved hybrid algorithm not only improves the parsing and sampling efficiency, but also realizes the optimal distribution of the sampling density function and improves the accuracy of the assessment index. The assessment indexs accurately quantify the risk values of three widely used cyber attacks.

Keywords: SCADA · Cyber attack · Risk assessment · Improved hybrid algorithm

1 Introduction

With the continued growth of electricity demand, the security of the power system is becoming increasingly important. In a power system, the application of the SCADA (Supervisory Control And Data Acquisition) is the most mature [1], which reliability ensures the security of the entire system.

Cyber attacks against SCADA occurs frequently in recent years, which have become a major factor affecting the security and reliability of power SCADA. Therefore, how to conduct the risk assessment of the power system has gradually been an urgent issue. We need to study an efficient risk assessment algorithm to accurately quantify the impact of cyber attacks on system reliability and predict potential threats to SCADA. The power system risk assessment algorithms can be roughly classified into two categories: parsing algorithms and Monte Carlo simulation algorithms [2].

H. Gao et al. (Eds.): CollaborateCom 2022, LNICST 461, pp. 3–17, 2022.
https://doi.org/10.1007/978-3-031-24386-8_1

The parsing algorithm obtains the random state of the system by fault enumeration and the probability of the random state by parsing calculation. The mathematical model of the parsing algorithm is accurate and the reliability indexes are highly precise. However, the number of system states to be analyzed by the parsing algorithm grows exponentially with the number of system components, which is difficult to apply to large-scale power system risk assessment.

The Monte Carlo simulation algorithm uses random sampling to obtain the status of each component in the power system, thereby determining the overall status of the power system and assessing system risk. The sampling number is independent of the size of the system. Therefore, it is particularly suitable for the risk assessment of large-scale power systems. However, the algorithm has a contradiction of calculation accuracy and sampling number [3,4]. The more precise the assessment index, the greater number of samples and the longer the calculation time required. We need to optimize the existing sampling algorithm to improve the convergence rate, which brings us great challenges.

In addition, most power SCADA risk assessments focus on the system itself, ignoring that cyber attacks are becoming a major factor affecting system security. Thus, we propose a novel power system risk assessment algorithm that takes into account cyber attacks. The major contributions of the work are four-fold:

- We propose an optimal sampling algorithm based on multiple integration models and variational problems, which realizes optimal sampling of the random state for improving the sampling efficiency and the indexes accuracy.
- To provide more efficient selection of parsing variables, we propose a selective parsing algorithm based on the projection variance, which overcomes the shortcomings of the parsing algorithm for the excessive analytical number.
- We combine the optimal sampling algorithm and the selective parsing algorithm into an improved hybrid algorithm for risk assessment of three attack types, which is the first attempt in the field. The improved hybrid algorithm combines the advantages of both algorithms.
- To assess the effectiveness of the improved hybrid algorithm, we conducted an error analysis of the risk assessment index and performed an experimental comparison on the UNSW-NB15 dataset. Experiments show that the algorithm can achieve better performance than other algorithms.

2 Related Work

The study of power system risk assessment algorithms has continuously been concerned by researchers.

Roslan [5] proposed sequential Monte Carlo (SMC) and non-sequential Monte Carlo (NSMC). They found that the SMC algorithm is more suitable to assess the distribution system. Zhang [6] proposed an improved SMC algorithm approach to substation connection risk assessment. Wu [7] adopted the SMC algorithm based on the minimal path sets to assess the risk of the distribution network. [8–10] proposed improved SMC algorithms to assess the risk of power systems respectively. However, the traditional Monte Carlo algorithm is memory-intensive, which leads to inefficiency.

Liu and Shen [11] used the improved important sampling algorithm, which meets the needs of assessment speed and accuracy. Bavajigari [12] presented the importance sampling algorithm to improve the computational efficiency of Monte Carlo sampling. Guo and Feng [13] calculated the security risk of the power system by the Latin hypercube sampling algorithm. Liu and Li [14] proposed a new algorithm combining Latin hypercube sampling (LHS) and Monte Carlo sequential simulation. The probability density functions adopted by these improved algorithms are superior to traditional sampling algorithms, but they are still far from the optimal distribution.

To overcome the shortcomings of the above algorithms, we propose a risk assessment algorithm that combines optimal sampling and selective parsing algorithms. The context of the assessment is that the power SCADA suffers cyber attacks. We aim to compare the performance of the risk indexes obtained by different algorithms and thus validate the superiority of the algorithm. The algorithm quantifies the impact of cyber attacks on power SCADA more accurately as well as has good engineering application value.

3 Preliminaries

3.1 SCADA Cyber Attack Types

Reports in [15] show an increasing number of security incidents and cyber attacks against SCADA in recent years. We have investigated the Repository of Industrial Security Incidents (RISI) [16] and SCADA cyber attacks that have occurred in the last 20 years [17]–[21] all over the world.

The three attacks that appear most frequently and bring us the biggest security challenges are *Analysis*, *DDoS*, and *Worm*. Specifically, *Analysis* contains the port scan, spam, and HyperText Mark-up Language (HTML) file penetrations. Attackers can use analysis tools to identify active ports and prepare for subsequent attacks. *DDoS* blocks the communication network by sending a large number of attack packets. Legitimate network packets are flooded with fake attack packets and can not reach the control center, while the network packets sent down from the control center can not be transmitted to the next layer of the network. *Worm* attacks Programmable Logic Controller (PLC) and other computers in the control center. Once the *Worm* infects the PLC, it can replicate itself to spread to other computers.

3.2 Traditional Risk Assessment Algorithm

To quantify the impact of three cyber attack types on the system, we take the 32 generators of the test system IEEE RTS-79 [22] as the example for the risk assessment. In the paper, we study the circuit breakers and generators of the power system as a whole object. Note that the circuit breakers and generators of the power system are treated as a whole object.

The forced outage rate (for) of a generator is the probability of an outage occurring when a component is forced out of operation immediately due to a

fault. for of the power system generator i corrected under conditions of cyber attack for_i.

The traditional Monte Carlo simulation algorithm samples each component and determines the component state. The combination gives the state of the entire system.

For the generator i, consider two states of normal operation (denoted by 1) and fault (denoted by 0):

$$x_i = \begin{cases} 0 & 0 \le U_i \le for_i \\ 1 & for_i < U_i \le 1, \end{cases} \tag{1}$$

where x_i is the state of generator i. U_i is a random number that obeys a uniform distribution U(0, 1) generated by a computer. By comparing U_i with for_i, the generator state x_i can be determined.

$LOLP$ (Loss of Load Probability) is the probability that the available capacity of a generation system will not be able to meet the annual maximum load demand of the system:

$$LOLP = \frac{1}{N} \sum_{i=1}^{N} F_{LOLP}(\vec{X}_i), \tag{2}$$

where N denotes the number of random states for the system. F_{LOLP} is the test function of $LOLP$. $\vec{X}_i$ is the system random state vector. When the system is in condition $\vec{X}_i$ without a load cut, then $F_{LOLP}(\vec{X}_i)$=0. Otherwise, $F_{LOLP}(\vec{X}_i)$=1.

$EDNS$ (Expected Demand Not Supplied) is the expected value of load demand power reduction due to generation capacity shortage in a given time range of the system, which is measured in MW:

$$EDNS = \frac{1}{N} \sum_{i=1}^{N} F_{EDNS}(\vec{X}_i), \tag{3}$$

F_{EDNS} is the test function of $EDND$. $F_{EDNS}(\vec{X}_i)$ represents the active power of the system in the random state $\vec{X}_i$ in accordance with the cut-off power.

$LOLP$ and $EDNS$ are both risk assessment indexes. The smaller value of both, the lower the risk value of the system. The higher the risk value of cyber attacks means the greater the threat to power SCADA.

4 Improved Risk Assessment Algorithm

The disadvantages of traditional sampling algorithms are low sampling efficiency and slow convergence. As a useful supplement to the sampling algorithm, the parsing algorithm can speed up the convergence rate. However, the number of parsing algorithm is excessive, increasing exponentially with the number of system components. We improve the two algorithms and combine them into an improved hybrid algorithm, as shown in Fig. 1.

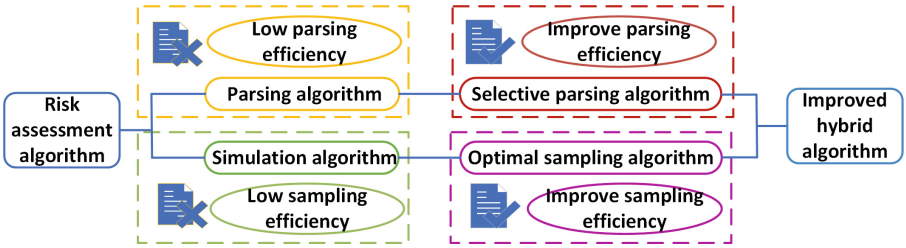

Fig. 1. The framework of the improved hybrid algorithm.

4.1 Multiple Integral Models for Risk Assessment

To reduce the variance of the test function, we need to optimize the probability distribution function of the system state variables. Rewrite the test function in the form of a random number vector $\vec{x}$ as the independent variable.

$$F(\vec{X}) = H(\vec{x}), \tag{4}$$

The element x_i of the vector $\vec{x}$ is a continuous variable.

$$R = E[F(\vec{X})] = E[H(\vec{x})] = \int_{\Omega} H(\vec{x}) d\vec{x}, \tag{5}$$

where R is the risk assessment index. Ω is an n-dimensional hypercube surrounded by planes $x_1 = 0$, $x_1 = 1$, $x_2 = 0$, $x_2 = 1$, $...x_n = 0$, $x_n = 1$. $d\vec{x} = dx_1 dx_2 ... dx_n$. We transform the power SCADA risk assessment problem into a multiple integral model.

4.2 Optimal Sampling Algorithm

Optimal Sampling Density Function. According to [11]–[14], reducing the variance $V\{H(\vec{x})\}$ of the test function can improve sampling efficiency and computational speed. Calculate the estimated value $\widehat{R}$ of the risk assessment index.

$$\widehat{R} = \frac{1}{N} \sum_{k=1}^{N} H'(\bar{x}_k), \tag{6}$$

$p(\vec{x})$ is the probability density function.
The probability density function of sample $\vec{x_k}$ is $p(\vec{x_k})$.

$$H'(\vec{x}) = \frac{H(\vec{x})}{p(\vec{x})}, \tag{7}$$

$H'(\vec{x})$ is the corresponding test function of sample $\vec{x_k}$. $\vec{x}$ takes continuous values in the integration region.

$$V\{H'(\vec{x})\} = \int_{\Omega} \left[H^2(\vec{x})/p(\vec{x})\right] d\vec{x} - \left[\int_{\Omega} H(\vec{x}) d\vec{x}\right]^2, \tag{8}$$

$V\{H'(\vec{x})\}$ is the variance of the test function. For the Eq. (8), $\left[\int_{\Omega} H(\vec{x})d\vec{x}\right]^2$ is a constant. When $J[p(\vec{x})] = \int_{\Omega} H^2(\vec{x})/p(\bar{x})d\bar{x}$ gets the minimum value, $V\{H'(\vec{x})\}$ gets the minimum value. Consider the independence of component states in power systems.

$$p(\vec{x}) = p_1(x_1)\, p_2(x_2) \cdots p_n(x_n) = \prod_{i=1}^{n} p_i(x_i). \tag{9}$$

The problem of minimizing the variance of the test function is transformed into a variational problem $J = \min\{J[p]\}$.

$$\begin{cases} J[p] = \int_{\Omega} \left[H^2(\bar{x})/\prod_{i=1}^{n} p_i(x_i)\right] d\bar{x} \\ \int_0^1 p_i(x_i)\, dx_i = 1 (i = 1, 2, \ldots, n). \end{cases} \tag{10}$$

According to the variational principle, the optimal edge distribution density of the i_{th} $(i = 1, 2, ...n)$ dimension is:

$$p_i(x_i) = \frac{\sqrt{\int_{\Omega_i} \left[H^2(\bar{x})/\prod_{\substack{j=i \\ j\neq i}}^{n} p_j(x_j)\right] d\vec{x}_{(i)}}}{\int_0^1 \sqrt{\int_{\Omega_i} \left[H^2(\vec{x})/\prod_{\substack{j=1 \\ j\neq i}}^{n} p_i(x_i)\right] d\bar{x}_{(i)} dx_i}}, \tag{11}$$

$d\vec{x}_{(i)} = dx_1 dx_2 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$. Ω_i is a subspace surrounded by planes $x_1 = 0$, $x_1 = 1$, $x_2 = 0$, $x_2 = 1$, $...x_{i-1} = 0$, $x_{i-1} = 1$, $x_{i+1} = 0$, $x_{i+1} = 1$,$...x_n = 0$, $x_n = 1$.

Optimal Sampling Algorithm. We set $I_i(x_i)$:

$$I_i(x_i) = \int_{Q_i} \left[H^2(\vec{x})/\prod_{\substack{j=1 \\ j\neq i}}^{n} p_j(x_j)\right] d\vec{x}_{(i)}, \tag{12}$$

The piecewise function $I_i(x_i)$ is constant in subintervals $[0, for_i)$ and $[for_i, 0]$:

$$I_i(x_i) = \begin{cases} I_{i1} & x_i \in [0, for_i) \\ I_{i2} & x_i \in [for_i, 1] \end{cases}, \tag{13}$$

Calculate the estimated value $\widehat{I_{i1}}$, $\widehat{I_{i2}}$ of I_{i1} and I_{i2}:

$$\hat{I}_{i1} = \frac{1}{N_1} \sum_{k=1}^{N_1} \left[\frac{H(\vec{x}_k)}{\prod_{\substack{j=1 \\ j\neq i}}^{n} p_j(x_j)}\right]^2 \Bigg|_{x_i \in (0, for_i)}, \tag{14}$$

$$\hat{I}_{i2} = \frac{1}{N_2} \sum_{k=1}^{N_2} \left[\frac{H(\bar{x}_k)}{\prod_{\substack{j=1 \\ j \neq i}}^{n} p_j(x_j)} \right]^2 \Bigg|_{x_i \in (for_i, 1)}, \tag{15}$$

N_1,N_2 are the number of samples that satisfy the two subintervals respectively. The expression of the optimal sampling density function is:

$$p_i(x_i) = \begin{cases} p_{i1} & x_i \in [0, for_i) \\ p_{i2} & x_i \in [for_i, 1] \end{cases}, \tag{16}$$

$$p_{i1} \approx \frac{\sqrt{\hat{I}_{i1}}}{for_i \sqrt{\hat{I}_{i1}} + (1 - for_i) \sqrt{\hat{I}_{i1}}}, \tag{17}$$

$$p_{i2} \approx \frac{\sqrt{\hat{I}_{i2}}}{for_i \sqrt{\hat{I}_{i2}} + (1 - for_i) \sqrt{\hat{I}_{i2}}}. \tag{18}$$

The whole algorithm is divided into two stages: pre-sampling and formal sampling. The purpose of pre-sampling is to obtain the optimal density function for each sub-interval by iterative calculation. Then, we sample the random states of the system according to the optimal density function during formal sampling by Eq. (17) and (18). Finally, the risk index of the power SCADA is assessed by Eq. (5).

4.3 Selective Parsing Algorithm

The selective parsing algorithm uses the projected variance to quantify the effect of the randomness of the variables on the variance of the test function. The variables with high impact are selected for parsing to effectively reduce the variance.

We set $K_i(x_i)$:

$$K_i(x_i) = \frac{\int_{\Omega_i} H(\vec{x}) d\vec{x}_{(i)}}{p_i(x_i)}. \tag{19}$$

Transform Eq. (6):

$$R = \int_0^1 dx_i \int_{\Omega_i} H(\vec{x}) d\vec{x}_{(i)} = \int_0^1 K_i(x_i) p_i(x_i) dx_i, \tag{20}$$

$$\hat{R}_i = \frac{1}{N} \sum_{j=1}^{N} K_i(x_{ij}). \tag{21}$$

The accuracy of $\widehat{R}$ depends on the variance of $K_i(x_{ij})$:

$$V\{K_i\} = \int_0^1 \{K_i(x_i) - E[K_i(x_i)]\}^2 p_i(x_i) dx_i, \tag{22}$$

$V\{K_i\}$ represents the effect of the randomness for the variable x_i on the variance of the risk assessment index, called the projected variance of x_i.

Analogous to $I_i(x_i)$, the piecewise function $K_i(x_{ij})$ is constant in subintervals $[0, for_i)$ and $[for_i, 0]$:

$$K_i(x_i) = \begin{cases} K_{i1} & x_i \in [0, for_i) \\ K_{i2} & x_i \in [for_i, 1] \end{cases}. \tag{23}$$

Calculate the estimated value $\widehat{K_{i1}}$, $\widehat{K_{i2}}$:

$$\hat{K}_{i1} = \frac{1}{N_1} \sum_{k=1}^{N_1} \left[\frac{H(\vec{x}_k)}{\prod_{j=1}^{n} p_j(x_j)} \right] \Bigg|_{x_i \in (0, for_i)}, \tag{24}$$

$$\hat{K}_{i2} = \frac{1}{N_2} \sum_{k=1}^{N_2} \left[\frac{H(\vec{x}_k)}{\prod_{j=1}^{n} p_j(x_j)} \right] \Bigg|_{x_i \in (0, for_i)}, \tag{25}$$

The projected variance of x_i is:

$$V\{K_i\} \approx \left\{ \hat{K}_{i1} - E[K_i(x_i)] \right\}^2 for_i p_{i1} + \left\{ \hat{K}_{i2} - E[K_i(x_i)] \right\}^2 (1 - for_i) p_{i2}, \tag{26}$$

$$E[K_i(x_i)] \approx for_i p_{i1} \hat{K}_{i1} + (1 - for_i r_i) p_{i2} \hat{K}_{i2}. \tag{27}$$

Similar to the optimal sampling algorithm, the selective parsing algorithm is divided into two stages: pre-sampling and formal sampling. Pre-sampling calculates the projection variances and arranges them in order of magnitude, which is used to select the parsing variables. The optimal set of parsing variables is determined in this order. The optimal parsing variables are analyzed in the formal sampling stage. We can improve parsing efficiency by the selective parsing algorithm.

4.4 Improved Hybrid Algorithm

We combine optimal sampling with selective parsing algorithms and propose an improved hybrid algorithm. The optimal sampling algorithm improves the efficiency of sampling calculation by optimizing the sampling density function. The selective parsing algorithm improves the efficiency of parsing calculation by optimizing parsing variables.

Figure 2 shows the flow chart of the proposed algorithm. First, we enter the system state, set the pre-sampling iterations and the sampling number to set the initial sampling density to 1. Then iteratively calculate the optimal density function for each dimension and interval in pre-sampling 1. Calculate the projection variance of variables according to the optimal density function and rank them in order of magnitude to determine the optimal parsing variables in pre-sampling 2. Next, sample the random states of the simulated variables

according to the optimal density function and enumerate the states of the parsing variables to obtain the system states in the formal sampling until sampling completes. Finally, count risk assessment indicators, test function variances and coefficients of variance, and output the assessment results.

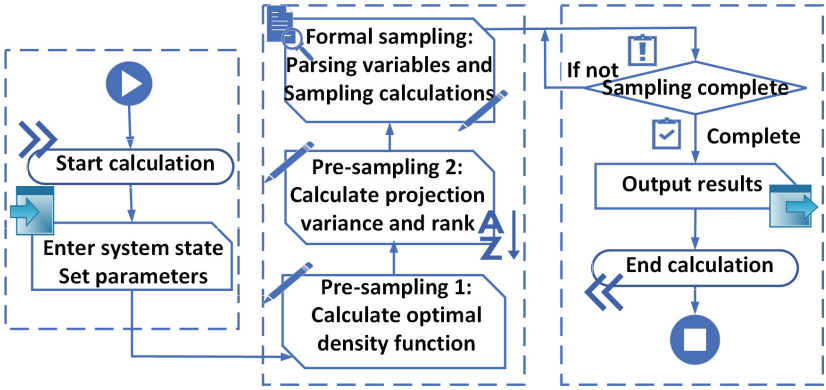

Fig. 2. Flow chart of the improved hybrid algorithm.

5 Example Analysis

5.1 Optimal Sampling Algorithm Assessment Results

To verify the sampling efficiency of the optimal sampling algorithm, we take 32 generators of IEEE-RTS79 as the object.

We conduct an error analysis of the reliability assessment indexes. Figure 3 shows β_{LOLP}, β_{EDNS}, V_{LOLP}, V_{EDNS} under different sampling number for four sampling algorithms. β_{LOLP}, β_{EDNS} is the variance of $LOLP$ and $EDNS$. V_{LOLP}, V_{EDNS} are the variances of the test function for $LOLP$ and $EDNS$. The pre-sampling of the optimal algorithm consists of two iterations of the calculation, each with 2000 samples. The number of samples starts from 6000.

It can be seen that the values of β_{LOLP}, β_{EDNS} of four sampling algorithms decrease as the number of samples increases. It means that the accuracy of the reliability index increases with the number of samples. The curves smooth out when the number of sampling reaches 20,000. At this point, the four sampling algorithms have the highest sampling efficiency and the most accurate calculation accuracy. Tests have shown that a sampling density function with minimal variance is obtained after 2 iterations for the optimal sampling algorithm.

As can be seen from the bar charts, the latter three improvements all reduce the sampling variance to some extent compared to the traditional algorithm. In particular, the optimal sampling algorithm has the largest reduction. It not only has the smallest variance of the test function under the same sampling number, but also the variance of the test function tends to decrease with each iteration. This proves that the index accuracy and sampling efficiency of the optimal sampling algorithm are the highest among the four algorithms.

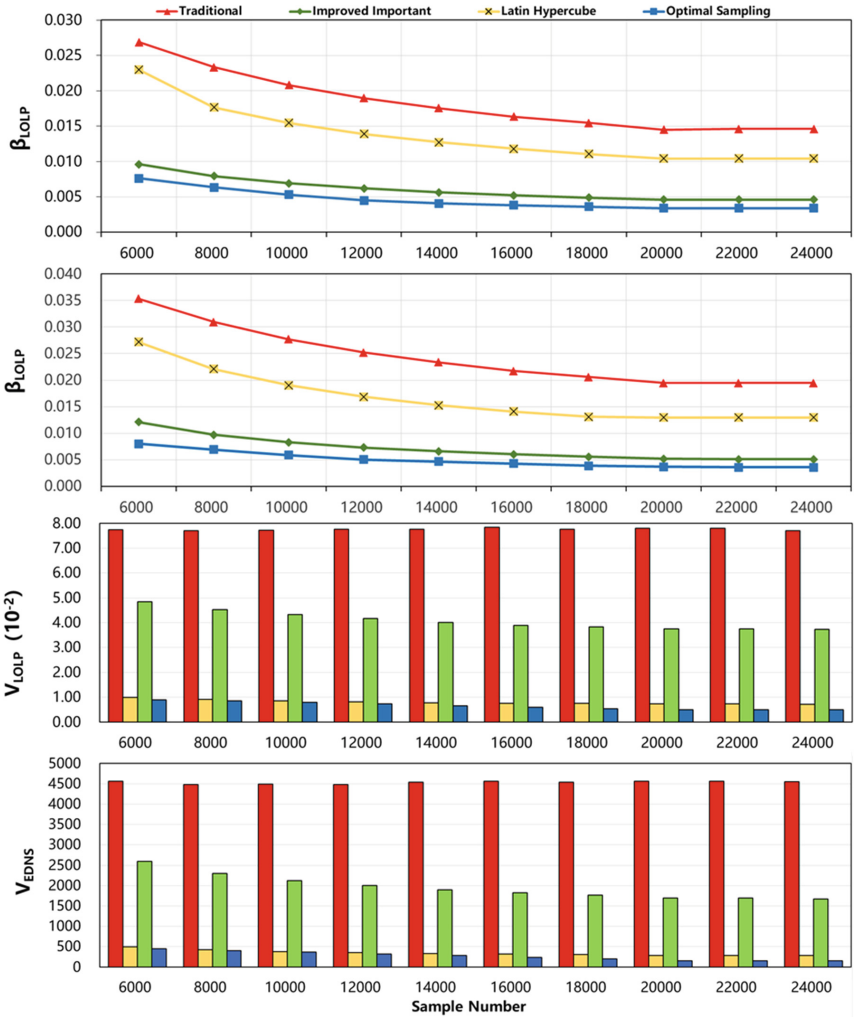

Fig. 3. β_{LOLP}, β_{EDNS}, V_{LOLP}, V_{EDNS} for four algorithms

5.2 Improved Hybrid Algorithm Assessment Results

To verify the performance of the improved hybrid algorithm, we assessed the selective parsing algorithm in combination with traditional and optimal sampling algorithms respectively. Based on the conclusion in Sect. 6.1, we selected a sample number of 20,000 for comparison. We consider the effect of the projection variance of the system state variables on reliability, which takes 32 generators of IEEE-RTS79 as the parsing variables.

Our experiment is divided into three stages.

In the first stage, we combine selective parsing with the traditional sampling algorithm to obtain an improved traditional algorithm. The pre-sampling is carried out according to the traditional sampling algorithm, which consists of two iterations of the calculation, each with 2000 samples. The projection variance of the variables is calculated from the obtained sampling density function by the selective parsing algorithm and the 32 generators are reordered according to the magnitude of the projection variance. In the formal sampling process, the reordered parsed variables are parsed. The variance of the test function is calculated to obtain the results of the improved traditional algorithm.

In the second stage, we combine selective parsing with the optimal sampling algorithm to obtain the improved hybrid algorithm. The traditional algorithm of the first stage is replaced with the optimal sampling algorithm. Repeat the steps of the first stage and obtain the results of the improved hybrid algorithm.

In the third stage, the performance of the improved algorithms are verified by comparing the variance of the test functions obtained in Sect. 5.1.

Figure 4 shows the comparison results of four algorithms. The abscissa is the generator number reordered according to the projected variance. The smaller the number, the larger the projected variance. We can conclude that V_{LOLP} and V_{EDNS} decrease as the projection variance increases. It means that parsing variables with larger projected variances have a greater impact on the variance of the test function. The greater the projection variance of the parsing variables, the more efficient the parsing of the selective parsing algorithm. Selecting parsing variables with large projection variance for parsing can reduce the variance of the test function and improve the parsing efficiency. The improved hybrid algorithm further improves the performance of the optimal sampling algorithm, which has the highest sampling efficiency in Sect. 5.1.

The number of samples is set to 20,000. We compare the assessment results of the five algorithms in Table 1. Note that $Time(/\mathrm{s})$ is the parsing time. From Table 1, compared to the other algorithms, we can conclude that β_{LOLP}, β_{EDNS}, V_{LOLP}, V_{EDNS} of the improved hybrid algorithm are minimum, which means that the improved hybrid algorithm has the highest sampling efficiency and the assessment indexes obtained are the most accurate. In addition, the improved hybrid algorithm greatly reduces the parsing time of algorithms due to the selective parsing algorithm's improved parsing efficiency. The parsing time improved 94.545% compared to the traditional algorithm. It is of great significance in the reliability assessment of modern large-scale power systems.

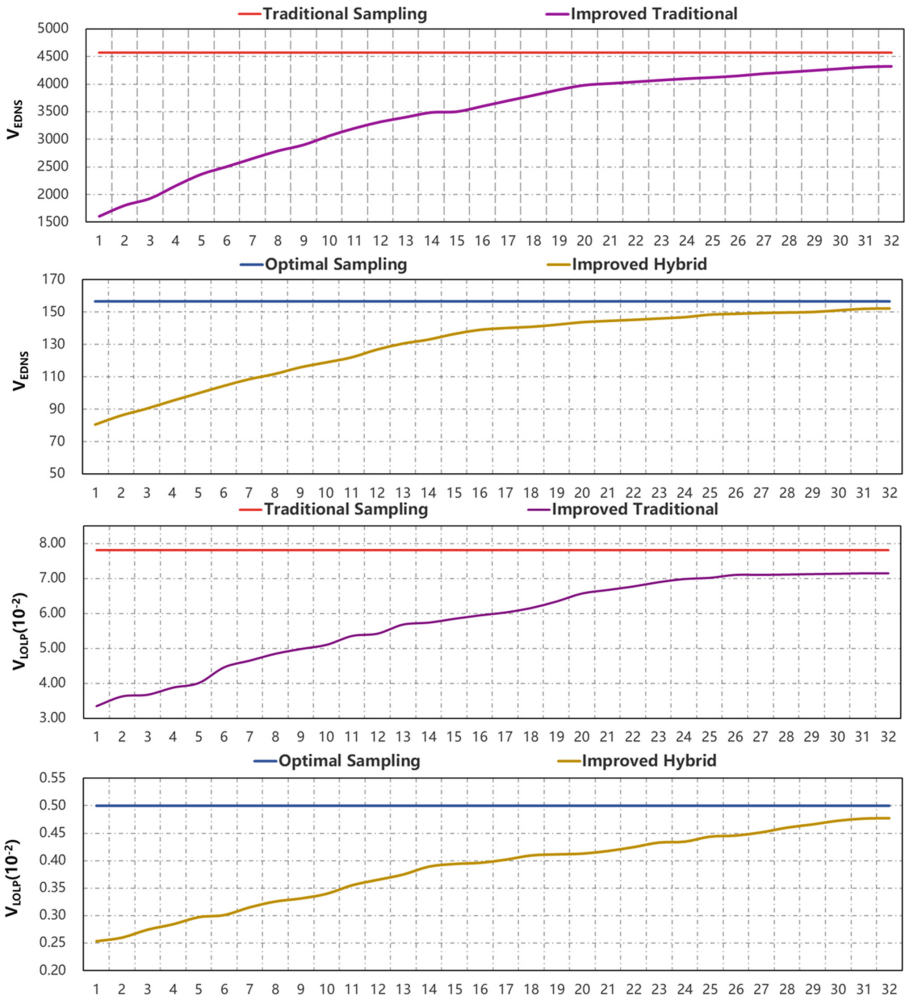

Fig. 4. V_{EDNS}, V_{LOLP} for improved algorithms

Table 1. The assessment results of five algorithms

Algorithm	$LOLP(10^{-2})$	$V_{LOLP}(10^{-2})$	$\beta_{LOLP}(10^{-2})$	$EDNS$	V_{EDNS}	$\beta_{EDNS}(10^{-2})$	Time
Traditional	8.43019	7.81100	1.45035	15.2672	4568.45	1.94486	165
Improved Important	8.49715	0.72946	0.45807	15.0129	289.65	0.52224	21
Latin Hypercube	8.50148	3.74545	1.04238	15.1606	1692.56	1.29878	132
Optimal Sampling	8.32678	0.50213	0.003406	14.5367	156.483	0.00369212	17
Improved Hybrid	**8.30146**	**0.32541**	**0.001534**	**14.5246**	**102.152**	**0.00161354**	**9**

Table 2. $LOLP(10^{-2})$ for cyber attack

Type	Traditional	Improved Important	Latin Hypercube	Optimal Sampling	Improved Hybrid
Primary value	8.4302	8.4972	8.5015	8.3268	**8.3015**
Worms	16.1869	16.3136	16.3459	15.8672	**15.9496**
DDoS	15.3428	15.4124	15.4423	15.1454	**15.0887**
Analysis	16.6942	16.8568	16.8154	16.5478	**16.4983**

Table 3. $EDNS$(/MW) for cyber attack

Type	Traditional	Improved Important	Latin Hypercube	Optimal Sampling	Improved Hybrid
Primary value	15.2672	15.0129	15.1606	14.5367	**14.5246**
Worms	30.0572	29.5787	29.8675	28.6785	**28.6033**
DDoS	31.5478	31.1878	31.4865	30.1853	**30.1502**
Analysis	33.9645	33.3752	33.7457	32.3458	**32.3071**

5.3 Risk Assessment for Cyber Attack

The UNSW-NB15 dataset [23] is used to generate attacks for evaluation. $LOLP$ and $EDNS$ are obtained for cyber attack in Table 2 and Table 3. Note that the primary values for the assessment index are obtained without cyber attacks. As a result of the five algorithms calculations, we can conclude that $LOLP$ and $EDNS$ values of *Analysis* attack are the largest among the three attack types. Through our error analysis it is clear that although the five algorithms produce consistent conclusions, the other four are not sufficiently precise in their indexes. To maximize the accuracy of the calculation, we have adopted the assessment indexes of the improved hybrid algorithm calculation. *Analysis* attack has the largest $LOLP$ and $EDNS$ values among the three attack types, meaning it has the greatest impact on the reliability of the system. The value of $LOLP$ increases 98.74% and $EDNS$ increases 122.43% compared to the primary value. For *Worms* attack, the value of $LOLP$ increases 92.13% and $EDNS$ increases 96.93%. For *DDoS* attack, the value of $LOLP$ increases 81.76% and $EDNS$ increases 107.58%. As a result of our precise assessment, the risk value of *Analysis* attack is the maximum, meaning it is the most dangerous for the power SCADA.

6 Conclusion

In this paper, we first propose an optimal sampling algorithm based on multiple integration models and variational problems, which improves sampling efficiency and indexes accuracy. Besides, a selective parsing algorithm based on the projection variance is proposed to provide a more efficient selection of parsing variables for improving parsing efficiency. Then, we combine the two improved algorithms to form an improved hybrid algorithm for risk assessment of three attack

types. After error analysis and experimental comparisons, we can confirm that the improved hybrid algorithm outperforms the traditional and existing algorithms. The assessment results accurately quantify the impact of cyber attacks on SCADA security, which show that the *Analysis* attack has the greatest risk value. It is extremely well predicted that *Analysis* attack is the greatest threat to power SCADA, providing insights into the establishment of the power system security enhancement strategies.

References

1. Darshana, U., et al.: An efficient key management and multi-layered security framework for SCADA systems. IEEE Trans. Netw. Serv. Manage. **19**(1), 642–660 (2021)
2. He, Hailei., et al.: Reliability evaluation based on modified latin hypercube sampling and minimum load-cutting method. In: 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT). IEEE, 2015
3. Gonzalez-Fernandez, R.A., Leite da Silva, A.M.: Reliability assessment of time-dependent systems via sequential cross-entropy Monte Carlo simulation. IEEE Trans. Power Syst. **26**(4), 2381–2389 (2011)
4. Bie, Z., Wang, X.: The application of Monte Carlo method to reliability evaluation of power systems. Autom. Electr. Power Syst. **21**(6), 68–75 (1997)
5. Roslan, N.N.R.B., Fauzi, N.F.B.M., Ridzuan, M.I.M.: Sequential and nonsequential monte carlo in assessing reliability performance of distribution network. In: 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE). IEEE (2020)
6. Zhang, K., et al.: Improved sequential monte carlo method approach to substation connection reliability assessment. In: 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). IEEE (2020)
7. Wu, L., et al.: Reliability assessment of AC/DC hybrid distribution network based on sequential monte carlo method. In: 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE). IEEE (2020)
8. Weibo, L., et al.: Risk assessment technology of ship power system based on improved time series algorithm. In: 2020 7th International Conference on Information Science and Control Engineering (ICISCE). IEEE (2020)
9. Li, L., et al.: Risk assessment for renewable energy penetrated power systems considering battery and hydrogen storage systems. In: 2021 Power System and Green Energy Conference (PSGEC). IEEE (2021)
10. Tang, S., Liu, Z., Wang, L.: Power system reliability analysis considering external and insider attacks on the SCADA system. In: 2020 IEEE/PES Transmission and Distribution Conference and Exposition (T&D). IEEE (2020)
11. Liu, J., Shen, H., Yang, F.: Reliability evaluation of distribution network power supply based on improved sampling monte carlo method. In: 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE). IEEE (2020)
12. Bavajigari, S.K.K., Singh, C.: Investigation of computational advantage of using importance sampling in monte carlo simulation. In: 2019 North American Power Symposium (NAPS). IEEE (2019)
13. Guo, J., et al.: Security risk assessment of power system based on latin hypercube sampling and daily peak load forecasting. In: 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). IEEE (2020)

14. Aibin, L., Wenyi, L.: Reliability evaluation of distribution network with distributed generation based on latin hypercube sequential sampling. In: 2020 3rd International Conference on Electron Device and Mechanical Engineering (ICEDME). IEEE (2020)
15. Miller, B., Rowe, D.: A survey SCADA of and critical infrastructure incidents. In: Proceedings of the 1st Annual Conference on Research in Information Technology (2012)
16. RISI-The Repository of Industrial Security Incidents, Apr 2020. [online] Available: http://www.risidata.com/
17. Pliatsios, D., et al.: A survey on SCADA systems: secure protocols, incidents, threats and tactics. IEEE Commun. Surv. Tutorials **22**(3), 1942–1976 (2020)
18. Moore, D., et al.: Inside the slammer worm. IEEE Secur. Priv. **1**(4), 33–39 (2003)
19. Levy, E.: The making of a spam zombie army. dissecting the Sobig worms. IEEE Secur. Priv. **1**(4), 58–59 (2003)
20. Langner, R.: Stuxnet: dissecting a cyberwarfare weapon. IEEE Secur. Priv. **9**(3), 49–51 (2011)
21. Samdarshi, R., Sinha, N., Tripathi, P.: A triple layer intrusion detection system for SCADA security of electric utility. In: 2015 annual IEEE India Conference (INDICON). IEEE (2015)
22. Stamp, J., McIntyre, A., Ricardson, B.: Reliability impacts from cyber attack on electric power systems. In: 2009 IEEE/PES Power Systems Conference and Exposition. IEEE (2009)
23. The UNSW-NB15 Dataset, Intelligent Security Group UNSW Canberra, June 2021 [online] Available: https://research.unsw.edu.au/projects/unsw-nb15-dataset