Appendix A: The Monotonicity of Poisson Binomial Distribution.

We state Lemma 1 which is used in the proof of Proposition 3.

LEMMA 1. *The cumulative distribution function of Poisson Binomial distribution is non-increasing w.r.t. success probability p_i for all $i = 1, \ldots, n$.*

PROOF. Let Y be a Poisson Binomial random variable with success probabilities $p_1, \ldots, p_n$. It suffices to show that the partial derivative of the cumulative distribution function of Poisson Binomial distribution w.r.t. p_i is nonpositive for all $i = 1, \ldots, n$. Without loss of generality, we concentrate on the nth Bernoulli random variable. The cumulative distribution function of Poisson Binomial distribution is given by:

$$F(y, p_1, \ldots, p_n) = \mathbb{P}(Y \le y) = \sum_{l=0}^{y} f(l, p_1, \ldots, p_n) = \sum_{l=0}^{y} \sum_{A \in \mathcal{B}_l(1,\ldots,n)} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j),$$

where $f(l, p_1, \ldots, p_n)$ denotes its probability mass function, i.e., the probability of l successes in n Bernoulli trials, and $\mathcal{B}_l(1, \ldots, n)$ denotes the set of all subsets of size l from $\{1, \ldots, n\}$. We can rewrite the probability mass function of Poisson Binomial distribution as follows:

$$\begin{aligned}
f(y, p_1, \ldots, p_n) &= \sum_{A \in \mathcal{B}_y(1,\ldots,n): n \in A} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j) + \sum_{A \in \mathcal{B}_y(1,\ldots,n): n \notin A} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j) \\
&= p_n \sum_{A \in \mathcal{B}_y(1,\ldots,n): n \in A} \prod_{i \in A: i \neq n} p_i \prod_{j \in A^c} (1 - p_j) + (1 - p_n) \sum_{A \in \mathcal{B}_y(1,\ldots,n): n \notin A} \prod_{i \in A} p_i \prod_{j \in A^c: j \neq n} (1 - p_j).
\end{aligned}$$

Let us now obtain the partial derivative of $f(y, p_1, \ldots, p_n)$ w.r.t. p_n. In fact, we have:

$$\begin{aligned}
\frac{\partial f(y, p_1, \ldots, p_n)}{\partial p_n} &= \sum_{A \in \mathcal{B}_y(1,\ldots,n): n \in A} \prod_{i \in A: i \neq n} p_i \prod_{j \in A^c} (1 - p_j) - \sum_{A \in \mathcal{B}_y(1,\ldots,n): n \notin A} \prod_{i \in A} p_i \prod_{j \in A^c: j \neq n} (1 - p_j) \\
&= \sum_{A \in \mathcal{B}_{y-1}(1,\ldots,n-1)} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j) - \sum_{A \in \mathcal{B}_y(1,\ldots,n-1)} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j).
\end{aligned}$$

Consider the quantity $H^{y-1} := \sum_{A \in \mathcal{B}_{y-1}(1,\ldots,n-1)} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j)$. The first term in the last equality follows from the fact that the index n indeed belongs to set $\mathcal{B}_y(1, \ldots, n)$, but is not used in any of the multiplication operations. This is equivalent to the selection of $y - 1$ many elements from $\{1, \ldots, n-1\}$. Similarly, consider the quantity $H^y := \sum_{A \in \mathcal{B}_y(1,\ldots,n-1)} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j)$. The second term in the last equality is due to the fact that the index n does not belong to set $\mathcal{B}_y(1, \ldots, n)$, and is not used in any of the multiplication operations. This is equivalent to the selection of y many elements from $\{1, \ldots, n-1\}$. Thus, the partial derivative of $f(y, p_1, \ldots, p_n)$ w.r.t. p_n is given by:

$$\frac{\partial f(y, p_1, \ldots, p_n)}{\partial p_n} = \begin{cases} -H^0 & \text{if } y = 0, \\ H^{y-1} - H^y & \text{if } 1 \le y \le n-1, \\ H^{n-1} & \text{if } y = n. \end{cases}$$

Then, it is easy to obtain the partial derivative of $F(y,p_1,\dots,p_n)$ w.r.t. p_n as follows:

$$\frac{\partial F(y,p_1,\dots,p_n)}{\partial p_n} = \begin{cases} -H^0 & \text{if } y=0, \\ -H^y & \text{if } 1\le y\le n-1, \\ 0 & \text{if } y=n. \end{cases}$$

Since $p_n \in [0,1]$, we clearly have $\frac{\partial F(y,p_1,\dots,p_n)}{\partial p_n} \le 0$ for $y=0,\dots,n$. This proves the property of monotonicity of Poisson Binomial distribution w.r.t. p_n. $\square$

Proof of Proposition 3. Consider any pair of maintenance decisions $v'=(w',z'), v''=(w'',z'')$ with the following property:

$$(h,t') \le (h,t'') \text{ for } (h,t')\in\mathcal{I}(v'),\ (h,t'')\in\mathcal{I}(v'') \text{ and } h\in\mathcal{H}'.$$

Let us first consider the set of generators prone to failure. As before, we let $\hat{\zeta}_{\mathcal{G}}(w')$ and $\hat{\zeta}_{\mathcal{G}}(w'')$ be the Poisson Binomial random variables with success probabilities $\{p'_i=\mathbb{P}(\xi_i\le m_i(w'));\ i\in\mathcal{G}\}$ and $\{p''_i=\mathbb{P}(\xi_i\le m_i(w''));\ i\in\mathcal{G}\}$, respectively. Clearly, the maintenance schedule under decision w'' is in a later period than the maintenance schedule under decision w', which implies that $m_i(w')\le m_i(w'')$ for $i\in\mathcal{G}'$. Then, we have $p'_i\le p''_i$. Secondly, we consider the set of transmission lines prone to failure. We let $\hat{\zeta}_{\mathcal{L}}(z')$ and $\hat{\zeta}_{\mathcal{L}}(z'')$ be the Poisson Binomial random variables with success probabilities $\{p'_{ij}=\mathbb{P}(\xi_{ij}\le m_{ij}(z'));\ (i,j)\in\mathcal{L}\}$ and $\{p''_{ij}=\mathbb{P}(\xi_i\le m_i(z''));\ i\in\mathcal{G}\}$, respectively. Similarly, we have $p'_{ij}\le p''_{ij}$ for $(i,j)\in\mathcal{L}'$.

By Lemma 1, we have $\mathbb{P}(\hat{\zeta}_{\mathcal{G}}(w')\le\rho_{\mathcal{G}})\ge\mathbb{P}(\hat{\zeta}_{\mathcal{G}}(w'')\le\rho_{\mathcal{G}})$ and $\mathbb{P}(\hat{\zeta}_{\mathcal{L}}(z')\le\rho_{\mathcal{L}})\ge\mathbb{P}(\hat{\zeta}_{\mathcal{L}}(z'')\le\rho_{\mathcal{L}})$. By using the independence of these random variables, we immediately have that $\mathcal{P}(v')\ge\mathcal{P}(v'')$. $\square$

Appendix B: Degradation Signal Modeling.

We model each degradation signal as a stochastic continuous process. For every $h\in\mathcal{H}$, we denote this process as $\mathcal{D}=\{D_h(t):t\ge 0\}$ with $D_h:\mathbb{R}\to\mathbb{R}$. given by

$$D_h(t)=\upsilon_h+\beta_h t+\sigma_h W(t), \tag{1}$$

where υ_h is the initial signal amplitude, β_h is the linear drift parameter and σ_h is the constant standard deviation of degradation signal of component h. The stochastic process $\mathcal{W}=\{W(t):t\ge 0\}$ is the standard Brownian motion with $W(0)=0$. We define the failure time of component h as the first passage time, i.e., $\xi_h=\min\{t\ge 0: D_h(t)\ge\Lambda\}$ for some predefined threshold Λ. We assume that the prior distribution of the initial signal amplitude is $\upsilon_h\sim\mathcal{N}(\mu_0,\kappa_0^2)$ and the prior distribution of the linear drift is $\beta_h\sim\mathcal{N}(\mu_1,\kappa_1^2)$ for every $h\in\mathcal{H}$. We let $D_h(t_h^i)$ be the degradation signal level of component h at time t_h^i. We define D_h^i as the increment between times t_h^i and t_h^{i-1}, given by $D_h^i=D_h(t_h^i)-D_h(t_h^{i-1})$ for $i=2,\dots,t_h^k$ with $D_h^1=D_h(t_h^1)$ where $t_h^k,\ k\in\mathbb{Z}_+$ is the random observation time of the degradation signal of component h. Given the observed data, we obtain the posterior distribution of the linear drift β_h (Proposition 2 by Gebraeel et al. (2005)), which is used to estimate the RLD of component h as in Proposition 1.

PROPOSITION 1. *Given the observed signal increments D_h^i at time $i=t_h^1,\dots,t_h^k$ with prior parameters (υ_h,β_h), and the predefined failure threshold Λ, the posterior mean of the drift parameter of component h is given by:*

$$\mu_h' = \frac{(\kappa_1^2\sum_{i=1}^{t_h^k} D_h^i + \mu_1\sigma_h^2)(\kappa_0^2+\sigma_h^2 t_h^1) - \kappa_1^2(D_h^1\kappa_0^2+\mu_0\sigma_h^2 t_h^1)}{(\kappa_0^2+\sigma_h^2 t_h^1)(\kappa_1^2 t_h^k+\sigma_h^2)-\kappa_0^2\kappa_1^2 t_h^1}. \tag{2}$$

Then, the remaining lifetime of component h at time t_h^k follows the inverse Gaussian distribution $\mathcal{IG}(t+t_h^k|\mu,\lambda)$ with shape parameter $\mu=\frac{\Lambda-\sum_{i=1}^{t_h^k} D_h^i}{\mu_h'}$ and scale parameter $\lambda=\frac{(\Lambda-\sum_{i=1}^{t_h^k} D_h^i)^2}{\sigma_h^2}$.

We generate a dataset consisting of unique degradation signals due to the lack of publicly available data to estimate the parameters of the prior distributions of υ_h and β_h for $h\in\mathcal{H}$. In power systems, it is realistic to assume that generators are more likely to fail than transmission lines (see, for example, Papavasiliou et al. (2015)). Thus, we follow this assumption with our dataset. For simplicity, we assume that the variance of υ_h and β_h are indeed known and held constant over the planning horizon for $h\in\mathcal{H}$. Therefore, we are only interested in estimating the prior mean of υ_h and β_h, denoted by μ_0 and μ_1, respectively. First, we focus on estimating μ_0 and μ_1 among the set of generators. For that purpose, we generate 100 unique degradation signals. Let us label these degradation signals with an index j where $j=1,\dots,100$. We assume that degradation signal j has the functional form (1) with $\upsilon_j\sim\mathcal{N}(20,10^2)$ and $\beta_j\sim\mathcal{N}(5,0.3^2)$ and $\sigma_j=3$ for $j=1,\dots,100$. The degradation signal threshold Λ is set to 100. We observe degradation signal j at discrete time points until a failure time $\xi_j=\{t: D_j(t)\geq 100,\ t\geq 0\}$ for $j=1,\dots,100$. We remind the reader that D_j^i is defined as the increment of degradation signals between times t_j^i and t_j^{i-1} for $i=2,\dots,\xi_j$ where $D_j^1=D_j(1)$, for $j=1,\dots,100$. We find the point estimate of μ_0 with $\sum_{j=1}^{100} D_j^1/100$. To obtain the point estimate of μ_1, we first compute the prior mean estimate of β_j as $\hat{\mu}_j=(\sum_{i=1}^{\xi_j} D_j^i - D_j^1)/\xi_j$ for $j=1,\dots,100$. Then, we find the point estimate of μ_1 with $\sum_{j=1}^{100}\hat{\mu}_j/100$. Eventually, we obtain the prior mean estimate among the set of generators. Secondly, to estimate μ_0 and μ_1 among the set of transmission lines, we follow a similar procedure after generating 100 unique degradation signals with $\upsilon_j\sim\mathcal{N}(15,5^2)$ and $\beta_j\sim\mathcal{N}(3,0.3^2)$ and $\sigma_j=1$ for $j=1,\dots,100$. Finally, we obtain the prior mean estimates of the stochastic parameters υ_h and β_h of the degradation signal model for every $h\in\mathcal{H}$.

Next, we obtain the posterior distribution of the unknown parameters of υ_h and β_h for $h\in\mathcal{H}$ with a Bayesian approach given the recently observed condition-based information. For that purpose, we generate 100 unique degradation signals with a random initial signal amplitude. For the sake of easier modeling, we assume that these degradation signals were observed at some random discrete times. We further assume that random observation time t_h^k for component h follows a uniform distribution on $[1,(\Lambda-\mu_0)/(\mu_1+3\kappa_1)]$. This assumption implies that degradation signal for component $h \in \mathcal{H}$ was observed when it had been drastically degrading with a gradual linear drift. Under these assumptions, we obtain the posterior mean of the drift parameter β_h of form (2), which easily yields us to identify the RLD of each component $h \in \mathcal{H}$ (Proposition 1). Consequently, we select set $\mathcal{H}'$ by means of RLDs as discussed in Section 3.1.

Appendix C: Parallel Computing

In order to demonstrate the effect of parallelism within Algorithm 1, we solve the 9-bus instance with a scenario size of 1000 by using the exact reformulation of the joint chance-constraint. The results of our computational experiment are presented in Figure 1 w.r.t. different number of threads.

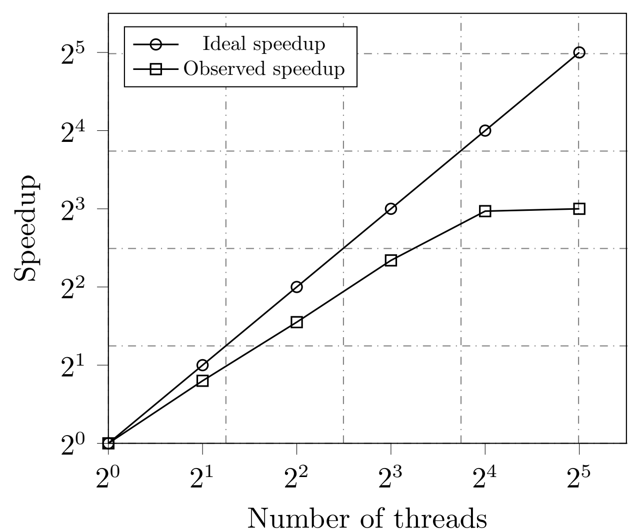

Figure 1 Speedup ratios with parallel computing.

Our empirical study indicates a sublinear growth in the parallel speedup ratios. We note that the data size of 9-bus instance is relatively small and the results are only representative, nevertheless, the utilization of the parallel computing becomes more apparent as the size of the problem increases.

References

Gebraeel NZ, Lawley MA, Li R, Ryan JK (2005) Residual-life distributions from component degradation signals: A bayesian approach. *IIE Transactions* 37(6):543–557.

Papavasiliou A, Oren SS, Rountree B (2015) Applying high performance computing to transmission-constrained stochastic unit commitment for renewable energy integration. *IEEE Transactions on Power Systems* 30(3):1109–1120.