A COMPUTATIONAL TOOL FOR A GENERAL MODEL OF OPERATION PROCESSES IN INDUSTRIAL SYSTEMS

Mohamed Salahuddin Habibullah, Ernest Lumanpauw

Institute Of High Performance Computing, Singapore
e-mail: mohdsh@ihpc.a-star.edu.sg, lumanpauwe@ihpc.a-star.edu.sg

Kołowrocki Krzysztof, Joanna Soszyńska

Gdynia Maritime University, Poland
e-mail: katmatkk@am.gdynia.pl, joannas@am.gdynia.pl

Ng Kien Ming

National University of Singapore, Singapore
e-mail: kienming@nus.edu.sg

ABSTRACT

The complexities of real industrial systems operation processes require computational methods that can analyze the large data and evaluate the behaviours of these systems. The use of methods such as Bayesian Network, Formal Safety Assessment and Statistical-Model based method were discussed as possibilities. Of which, a computational tool, based on the Semi-Markov model, was developed. This tool was then applied to analyze the behaviour of the operation processes of the oil transportation system in Dębogórze, Poland. The analyses showed that the computational solutions generated compared favourably well with some well-established analytical formulae, enabling possible extensions of the tool to include reliability and optimization evaluations to be explored.

1 INTRODUCTION

Many real industrial systems belong to the class of complex systems, resulting from the large number of components and interconnected parts, which are collectively assembled to define the operations and properties of the systems. Due to the complexity of such systems, it often causes the evaluation of the system reliability, availability and safety to become difficult. These complexities are multiplied in the case of large complex systems, where the determination of the exact reliability, availability and risk functions of the systems, leads to very complicated formulae, often useless for reliability practitioners to use. In real maritime systems, the focus area of this paper, some examples of such systems are in the piping transportation of water, gas and oil (Guze S. et al 2008) as well as in shipyard transportation using belt, rope conveyers and elevators (Blokus-R. et al 2008).

The difficulties associated with these large complex systems are further compounded when reliability optimization of these systems needs to be evaluated, with respect to their safety and costs. Due to the mathematical complexities of the current methods, such evaluations are often complicated and not possible to be performed by practitioners. In addition, the need to analyze these systems in their variable operation conditions as well as considering their changes in time reliability
structures and observing the components reliability characteristics, further complicates the issue. Furthermore, in handling such systems, the large datasets generated out of these systems, which needs to be analysed and processed, often necessitates the use of data mining tools and extensive compute-power, to speed up the computational processes. Thus, the availability of a computational tool that can model such large and complex maritime industrial systems operation processes, would indeed be valuable to practitioners and users.

In developing such computational tool, various approaches have been proposed in understanding the behaviour of maritime industrial systems operation processes. The 3 most commonly researched methods are namely the Bayesian Network (BN), the Formal Safety Assessment (FSA) and the Statistical Model-based method. Here, an overview of these 3 methods and its applicability in evaluating the behaviour of maritime industrial systems operation processes is discussed.

Bayesian Network (BN) (Pearl J. 1985) is a probabilistic model, whose nodes are used to represent variables and edges, in describing variables dependence with each another. It is a popular research approach to the maritime industry and it is often considered when analyzing the probability relationship between the different types of ships and accident rates. It is also a model that can be used to clearly represent the inter-relationship between subsystems and components of a system. In modelling maritime transportation systems, some variances of the BN have emerged. The Bayesian Belief Network (BBN) (Trucco P. 2008) was primarily developed for modelling Maritime Transport System (MTS), by taking into consideration ship owners, shipyards, port authorities, regulators and their mutual influences. By considering the case for the design of High Speed Craft (HSC), the study looked into the risk analysis associated with the quantification of Human and Organizational Factors (HFO). Integrated with the Fault-Tree analysis, the BBN managed to identify the probabilistic correlations between the basic events of collision and the model under HFO conditions, for collision in the open seas. Norrington (Norrington L. et al 2008) had also used the BBN model to conduct the reliability analysis of Search-And-Rescue (SAR) operations with the UK coastal guard coordination centres. Another variant is the Fuzzy-Bayesian Network (FBN) (Eleye-D 2008), applied in marine safety assessment, by integrating human elements into quantitative analysis. For this analysis, mass assignment theory was used as the bridge to connect the human factor and the probabilistic calculation.

Formal Safety Assessment (FSA) is a structured and systematic methodology, aimed at enhancing maritime safety by using risk analysis and cost benefit assessment. It is achieved by providing justifications for the proposed regulatory measures and allowing comparisons of the different measures to be made. This is in line with the basic philosophy of the FSA in that it is a tool that can be used to facilitate transparent decision-making process. FSA is also used to help in evaluating new regulations for maritime safety and protection of marine environment, with the aim of achieving a balance between technical and operational issues, which includes human, maritime safety, protection of marine environment and costs (IMO 2002). In its application to maritime systems operation processes, Ruud (Rudd S. et al 2008) has applied the FSA in developing risk-based rules and functional requirements, for systems and components in an offshore crane system. In studying the watertight integrity of hatchways of bulk carriers, Lee (Lee J. et al 2001) has also applied the FSA, which resulted in 18 hazards to be identified. This enabled 32 risk control measures to be devised in reducing the associated risks. Wang (Wang J. 2002) also explored the use of the FSA in maritime design. By considering both offshore and marine safety, the current practices as well as recent developments in safety assessment were illustrated. This was then applied to several maritime case studies, resulting in the relationship between offshore safety and formal ship safety assessment to be described. Others have also tried incorporating new models within the FSA. One such effort is by Hu (Hu S. et al 2007) who proposed a Model based on Relative Risk Assessment (MRRA) and used it to discuss the frequency and severity criteria affecting in ship navigation.
The third approach is the Statistical Model-based method, defined as a set of mathematical equations which describes the object of interest in terms of random variables and its associated probability distribution. From the statistical analysis point of view, in applying such methods, availability of real data is a necessity as these data are then used to extract features in relation to maritime safety, reliability, availability and risks. One such real data is from the Marine Accident Investigation Branch (MAIB) in UK. Using data from 1992 to 1999, Wang (Wang J. et al 2005) conducted a comprehensive statistical analysis of accidents involving fishing vessels, with the results presented in tables and graphs, showing the frequency and trends of the various accidents. Jin (Jin D. 2005) also conducted similar fishing vessels accidents analyses using data from northeastern USA but investigated further by modelling the accidents using logistics probability distribution. By considering factors such as weather and vessel size into the probability function, the paper provided a complete analysis of all causal factors of the fishing accidents and how much do they affect the ship safety, leading to the conclusion that the accidents probability is affected by the weather, vessel location, time of year, and vessel characteristics. Thus, it can be seen that the primary aim of these statistical models is to simulate, and then predict and optimize the system. One particular branch of statistical methods that has such capabilities is the Markov and Semi-Markov (Heyman D. et al 2003) methods. These methods have been used to estimate the corrosion rate as well as finding the failure probability of piping system (Vinod G. et al 2003), to study the patient population profile of a clinical trial, for arbitrary many patient classes, trial sites and start-times (Felli J. C. et al 2007) and to model in conjunction with Weibull distribution holding times to the actual power-plant operating data (Perman M. et al 1997).

In this paper, the Statistical Model-based method was chosen to provide a mean of understanding the behaviour of maritime industrial systems operation processes. This is due to the fact that the Statistical Model-based method is capable of handling mathematical complexities associated with maritime systems. The ability of this method to handle large datasets together with the use of data mining tools and extensive compute-power is also a plus. Also, recent breakthroughs in the use of Semi-Markov models on maritime transportation by Kołowrocki (Guze S, et al 2008, Blokus-R et al 2008, Kołowrocki K. et al 2008) have opened many possibilities. In his papers, he has analytically modelled the maritime transportation operations as a Semi-Markov process, with the stationary limiting probability that the operation will stay in each state, are computed based on the model. He then applied this to several real problems such as shipyard rope transportation system port oil transportation and ship operational process. Although generalized BN and FSA can possibly perform similar studies, the focus here is more towards leveraging the work done by Kołowrocki (Guze S, et al 2008, Blokus-R et al 2008, Kołowrocki K. et al 2008). Comparisons of the 3 methods with actual data, on its treatment of uncertainties, will be explored and published elsewhere.

Although the Semi-Markov model can be evaluated analytically, its solution is straightforward only for simple problems, with limited number of states considered. For large and complex systems such as in real maritime transportation, the continued use of analytical computation is tedious and cumbersome. Thus, having a Semi-Markov computational tool that can model industrial systems operation processes would indeed be advantageous, as the reusability of the codes would enable users to continuously improve and optimize the operation processes. In addition, the use of such computational tool would also ease the handling of the compute-intensive large data sets associated with maritime transportation and its use together with data mining techniques.

The rest of this paper is organized as follows. Section 2 describes the proposed general Semi-Markov model of industrial systems operation processes. This is then followed by Section 3, where the developed Semi-Markov computational tool, is discussed. In Section 4, the application of the computational tool for the case of port oil transportation system process is analyzed and described. Finally, in Section 5, the paper is concluded with outlines of future possible works that could be explored.
2 GENERAL MODEL OF INDUSTRIAL SYSTEMS OPERATION PROCESSES

The general model of industrial systems operation processes is formulated as a Semi-Markov process (Grabski F. 2002). The systems, during its operation processes, are taking $v, v \in N$ different operation states. Furthermore, $Z(t), t < 0, +\infty >$ is defined as the process with discrete operation states from the set of states, $Z = \{z_1, z_2, ..., z_v\}$, and conditional sojourn times, $\theta_{bl}$, at the operation states, $z_b$, when its next operation state is $z_l, bl = 1, 2, ..., v, b \neq l$. Based on the above assumptions, the general system operation process may be described by:

- The initial operation state probability vector:
  \[ p_b(0) = [p_1(0), p_2(0), ..., p_v(0)] \]  \hfill (1)
  where $p_b(0) = P(Z(0) = z_b)$ for $b = 1, 2, ..., v$

- The transition probability matrix:
  \[
  [p_{bl}]_{v,v} = \begin{bmatrix}
  p_{11} & p_{12} & \cdots & p_{1v} \\
  p_{21} & p_{22} & \cdots & p_{2v} \\
  \vdots & \vdots & \ddots & \vdots \\
  p_{v1} & p_{v2} & \cdots & p_{vv}
  \end{bmatrix}
  \]  \hfill (2)
  where $p_{bb} = 0$ for $b = 1, 2, ..., v$

- The conditional sojourn time distribution matrix:
  \[
  H_{bl}(t)_{1,v} = \begin{bmatrix}
  H_{11}(t) & H_{12}(t) & \cdots & H_{1v}(t) \\
  H_{21}(t) & H_{22}(t) & \cdots & H_{2v}(t) \\
  \vdots & \vdots & \ddots & \vdots \\
  H_{v1}(t) & H_{v2}(t) & \cdots & H_{vv}(t)
  \end{bmatrix}
  \]  \hfill (3)
  where $H_{bl}(t) = P(\theta_{bl} < t)$ for $b = 1, 2, ..., v, b \neq l$ and $H_{bb}(t) = 0$ for $b = 1, 2, ..., v$

- The mean value of the conditional sojourn time:
  \[
  M_{bl} = E[\theta_{bl}] = \int_0^\infty h_{bl}(t)dt
  \]  \hfill (4)
  where $h_{bl}(t) = \frac{d}{dt}[H_{bl}(t)]$ for $b = 1, 2, ..., v, b \neq l$

The use of the Semi-Markov model is mainly due to the scarcity of available data. In such scenarios, the Semi-Markov model is perhaps the best available option to analyze the problems, since it is built upon theories of probability and stochastic process, and hence justifying the use of the density function to get the mean, as in equation (4). It is also common practice that as more data becomes available, data mining methods such as regression, prediction, etc as well as simulation tools can be used to build a more robust analytical model of the problem. Thus, from the law of total probability,

- The distribution function of the unconditional sojourn time $\theta_b$ is given by:
\[ H_b(t) = \sum_{l=1}^{v} p_{bl} H_{bl}(t), \quad b = 1, 2, \ldots, v \]  
\[ (5) \]

- The mean value of the unconditional sojourn times is given by:
\[ M_b = E[\theta_b] = \sum_{l=1}^{v} p_{bl} M_{bl}, \quad b = 1, 2, \ldots, v \]  
\[ (6) \]

The limiting probability is one of the important characteristics of the process, it is given by:
\[ P_b = \lim_{t \to \infty} p_b(t) = \frac{\pi_b M_b}{\sum_{i=1}^{v} \pi_i M_i}, \quad b = 1, 2, \ldots, v \]  
\[ (7) \]

where the stationary distribution vector, \([\pi_b]_{1 \times v}\), is a vector which satisfies the following system of equations:
\[ [\pi_b] = [\pi_b] p_{bl}, \quad \sum_{i=1}^{v} \pi_i = 1 \]  
\[ (8) \]

Thus, having obtained the limiting probability values, the expected time spent in a particular operation state for sufficiently large operation time, \(\theta\), can be approximated by:
\[ E[\theta_b] = p_b \theta, \quad b = 1, 2, \ldots, v \]  
\[ (9) \]

### 3 COMPUTATIONAL TOOL FOR MODELLING INDUSTRIAL SYSTEMS OPERATIONS PROCESSES

Here, the computational tool for the general probabilistic model of industrial systems operation processes, described in the previous section, is presented. The basis on the development of the computational model is based on the number of states, \(v\), that the problem needs to handle. When the number of states, \(v\), is small, it is still possible to perform all the calculations analytically by hand, as undertaken by Kołowrocki (Guze S, et al 2008, Blokus-R et al 2008, Kołowrocki K. et al 2008). However, when the value of \(v\) is large, these calculations will be overly tedious and difficult, in particular the calculations associated with equation (8). Therefore, having a computational tool software to automate the tasks in Section 2 will undoubtedly assist in analyzing and solving the model, and hence the problem.

Due to the convenience in handling large matrices, the model is coded using a matrix-based numerical programming language. In performing this, various tools currently exists and have been explored namely Matlab, GNU Octave and R. In our implementation, the GNU Octave tool is employed since it is an open-source and free programming language. Furthermore, the code developed is mostly compatible with Matlab, a widely-used commercial software.
Figure 1. Block diagram of the general model of industrial systems operation processes.

Typically, the structure of a computer program is best described by its inputs, outputs and computation procedures or algorithms. In the case for general model of industrial systems operation processes considered, the inputs, outputs and model relationship can be explained by the block diagram illustrated in Figure 1. For brevity, since the model is based on the Semi-Markov process, the model is denoted as the Semi-Markov model.

From Figure 1, it can be seen that to generate the outputs, i.e. $M_b$, $[\pi_b]$, $p_b$, the adopted computer tool will take in the values of the conditional sojourn time, $\theta$, the transition probability matrix, $[p_{bl}]_{x\times x}$, and the conditional sojourn time distribution matrix, $[H_{bl}(t)]_{x\times x}$. If the Semi-Markov model is considered as a black box, then the former 3 values will act as inputs in generating the 3 outputs.

![Flowchart](image-url)

Figure 2. Flowchart of the Semi-Markov model computational procedures.

Figure 2 describes the details of the computational Semi-Markov model adopted in this paper. As shown in the flowchart, upon reading in the 3 inputs, the value of the integral, $M_{bl}$, needs to be evaluated. This evaluation is often difficult, especially if the density function of the conditional sojourn time distribution, $h_{bl}(t)$, is complex. Furthermore, some numerical programming languages such as Octave do not have the capabilities to handle such complex integrations. An alternative approach is to numerically approximate the mean value of $M_{bl}$ by generating sufficiently large...
random samples from the corresponding distribution density function and then averaging their values. In our implementation, 100,000 samples were generated. Once this is done, the value of $M_b$ was then evaluated.

The next stage is the most compute-intensive aspect of the calculations, which involves the evaluation of the stationary distribution vector, $[\pi_b]$. It can be seen from equation (8) that the evaluation of $[\pi_b]$ involves solving a series of linear equations, which is simple if the number of states is small. However, when the number of states considered in the problem is large, the solution process is rather complex and compute-intensive in nature. To overcome this, we propose the adoption of the following machine oriented approach. Our proposition involves letting $[P_{bl}]$ be a $v \times v$ irreducible transition probability matrix. Let us suppose $I$ is a $v \times v$ identity matrix and $ONE$ is a $v \times v$ matrix whose entries are all 1. This would lead to:

$$[\pi_b] = (1,...,1)(I - [p_{bl}] + ONE)^{-1}$$

(10)

The proof of the above proposition can be derived from equation (8), where upon simple rearrangement leads to:

$$[\pi_b] (I - [p_{bl}]) = 0$$

(11)

Furthermore, from the system of equations hypotheses, it is known that $[\pi_b]$ must sum to 1. Thus, from equation (10), it leads to:

$$[\pi_b] (I - [p_{bl}] + ONE) = [\pi_b] ONE = (1,...,1)$$

(12)

Furthermore, if we now assume that $(I - [p_{bl}] + ONE)$ has an inverse, thus solving for $[\pi_b]$ will yield equation (10) as desired. The adoption of this machine learning approach will assist in overcoming the compute-intensive nature of the calculations. Once the value of $[\pi_b]$ is obtained, as shown in the flow chart, the value of $p_b$ can then be evaluated using equation (7). This will then lead to the value of $E[\hat{\theta}_b]$ from equation (9) to be computed.

4 APPLICATION ON A PORT OIL TRANSPORTATION SYSTEM PROCESS

The proposed computational tool described, in the previous section, is now applied on the case of analyzing the operation processes of the port oil transportation system in Dębogórze, Poland (Guze S. 2008). As shown in Figure 3, the process involves the transportation of liquid cargo to Dębogórze terminal from the pier in the Port of Gdynia. In our analysis, the considered system is composed of 3 stages, namely, the pier, the 3 terminal parts A, B and C and the 3 linked piping subsystems, $S_1$, $S_2$ and $S_3$. The breakdown of the subsystems is as follows:

- Subsystem $S_1$: Consists of 2 identical pipelines, each composed of 178 elements. In each pipeline, there are 176 pipe segments and 2 valves.
- Subsystem $S_2$: Consists of 2 identical pipelines, each composed of 719 elements. In each pipeline, there are 717 pipe segments and 2 valves.
- Subsystem $S_3$: Consists of 2 types 1 pipeline and 1 type 2 pipelines, each composed of 362 elements. In each of the type 1 pipeline, there are 360 pipe segments and 2 valves. In each of the type 2 pipeline, there are 360 pipe segments and 2 valves.
At the pier, the unloading of tankers is performed in the Port of Gdynia. The pier is connected with the terminal Part A through the transportation subsystem $S_1$. In terminal Part A, there is a supporting station to fortify the tankers’ pumps, enabling the further transportation of the oil cargo to the terminal Part B through the subsystem $S_2$. The terminal Part B is further connected to the terminal Part C via the subsystem $S_3$. Finally, in the terminal Part C, the rail cisterns are unloaded with oil cargo products to be distributed to the rest of Poland.

**Figure 3.** Schematic view of the Port Oil Transportation System.

Based on the operation processes taking place at the abovementioned system, the process $Z(t)$, where $t \in (-\infty, +\infty)$, is defined, with the operation states as follows:

- $z_1$: Transportation of 2 different medium types from terminal Part B through Part C using two out of three pipelines in subsystem $S_3$.
- $z_2$: Transportation of 1 medium type from terminal Part C through Part B using one out of three pipelines in subsystem $S_3$.
- $z_3$: Transportation of 1 medium type from terminal Part B through Part A to the pier using 1 out of 2 pipelines in subsystem $S_2$ and 1 out of 2 pipelines in subsystem $S_1$.
- $z_4$: Transportation of 2 medium types from the pier through Parts A and B to Part C using both pipelines in subsystem $S_1$, both pipelines in subsystem $S_2$ and 2 out of 3 pipelines in subsystem $S_3$.
- $z_5$: Transportation of 1 medium type from the pier through Parts A and B to Part C using 1 out of 2 pipelines in subsystem $S_1$ and $S_2$ and 1 out of 3 pipelines in subsystem $S_3$.

Thus, using the 5 operation states described above, the computational procedures presented in Figure 2, can now be applied. In our analysis, due to the scarcity of historical data, the values of the transition probability matrix, $[P_{bl}]$ and the conditional sojourn time distribution matrix, $[H_{bl}(t)]$, were constructed by eliciting opinions from domain experts. This leads to the followings:

$$[H_{bl}(t)] = \begin{bmatrix} 0 & 0 & 0 & 1 - e^{-11484695t^2} & 0 \\ 0 & 0 & 1 - e^{-107737.1t^2} & 0 & 0 \\ 1 - e^{-37117.4t^2} & 0 & 0 & 0 \\ 0 & 1 - e^{-107737.1t^2} & 0 & 0 \\ 0 & 0 & 1 - e^{-29.1t^2} & 0 \end{bmatrix}$$
With the inputs given in equation (13), the values of the integral, $M_{bl}$, was then evaluated leading to the values of the unconditional sojourn time distribution function, $H_b(t)$, and its mean value $M_b$ for $b=1,2,\ldots,5$ to be given by:

$H_1(t) = 1 - \exp[-37117.4t^2] \quad M_1 = 0.005$

$H_2(t) = 1 - \exp[-19174.9t^2] \quad M_2 = 0.006$

$H_3(t) = 1 - 0.11 \cdot \exp[-148469.5t^2] - 0.89 \cdot \exp[-107737.1t^2] \quad M_3 = 0.003$

$H_4(t) = 1 - 0.5 \cdot \exp[-96963.4t^2] - 0.5 \cdot \exp[-29.1t^2] \quad M_4 = 0.001$

$H_5(t) = 1 - \exp[-29.1t^2] \quad M_5 = 0.164$

Here, the most compute-intensive aspect of the calculations, which involves evaluating the stationary distribution vector, $[\pi_b]$, is undertaken. Thus, using the computational procedures described in equation (10), leads to the values of the stationary distribution vector of the process to be given by:

$$[\pi_b] = \begin{bmatrix} 1/22 & 9/22 & 9/22 & 2/22 & 1/22 \end{bmatrix}$$

(14)

This enables the limiting probability, $P_b$ for $b=1,2,\ldots,5$ to be evaluated, yielding,

$$[P_1\ P_2\ P_3\ P_4\ P_5]=[0.018\ 0.228\ 0.095\ 0.007\ 0.652]$$

(15)

Thus, from the above computations, the expected time spent in a particular operation state, $E[\hat{\theta}_b]$ for $b=1,2,\ldots,5$, given $\Theta = 365$ days, were then evaluated.

$$\begin{bmatrix} E[\hat{\theta}_1] & E[\hat{\theta}_2] & E[\hat{\theta}_3] & E[\hat{\theta}_4] & E[\hat{\theta}_5] \end{bmatrix} = [6.6\ 83.2\ 34.7\ 2.6\ 238] \text{ days}$$

(16)

The computed values of $E[\hat{\theta}_b]$ were then compared with the analytical solutions evaluated by Kołowrocki (Guze S. 2008) showing very favourable comparisons. This provides a preliminary indication on the ability of the computational Semi-Markov model tool developed as well as its continued use to other maritime problems and possible extension to include the evaluation of reliability, availability and risk of the systems.

5 CONCLUSIONS
The paper has introduced a computational tool, based on the Semi-Markov model that can be used to analyze the general stochastic model of industrial systems operation processes. The application of the developed computational tool was then illustrated for the port oil transportation system. The results showed that the computational solutions matched well with the analytical calculations. This preliminary result from the developed computational tool showed the potential of the tools’ practical usefulness in other operation process evaluations, especially under changing structures and characteristics. In the long term, the aim is for this computational tool to be extended to incorporate reliability and availability calculations as well as optimization modules, using large-scale data from the maritime industrial systems operations processes. The block diagram of the proposed process workflow is shown in Figure 3 below. The implementation and results of the extended computational tool will be published elsewhere.

6 ACKNOWLEDGEMENTS

This paper describes part of the work in the Singapore-Poland Joint Research Project titled “Safety and Reliability of Complex Industrial Systems and Process” (SERC grant number 072 1340050) supported by grants from the Agency for Science, Technology and Research (A*STAR) of Singapore and the Ministry of Science and Higher Education (MSHE) of Poland.

REFERENCES