ABSTRACT

Process termination is a common strategy that is used to recover from deadlocks. However, terminating processes during their execution may affect and degrade the performance of the underlying system. The proposed solution in this paper is to select particular processes that can reduce the potential consequences of process termination in order to be terminated. A goal programming (GP) model is constructed to identify and select the best processes that can break a deadlock at lowest consequences of process termination. Several experimental tests are performed and the results showed that the proposed solution maintains the performance of the system during deadlock recovery compared to the other related methods.

Keywords

Goal Programming, Deadlock Recovery, Process Termination, Termination Cost.

1. INTRODUCTION

A deadlock in a computer system is a situation in which several processes are waiting infinitely to allocate some resources that are allocated by other waiting processes and cannot complete its tasks [1]. As a result, such situation can prevent the system from running and achieve its designated goals. Deadlock is an important issue that needs to be handled in computer systems, and in operating systems (OS) particularly [2-3]. Among the common strategies of deadlock handling, deadlock detection and recovery is accepted as a main feasible solution due to its applicable principles about resource allocation requirements of the processes [4-9].

Recovery from a deadlock can be performed when a deadlock occurs by terminating all or some deadlocked processes and releasing its allocated resources to be available for use [10]. However, aborting processes is found to be challenging because a great cost might result; utilizing of several allocated resources might be interrupted and part of computations and performed work might be wasted which negatively impacts the system performance [11]. Several desirable performance attributes for the system, such as throughput and resource utilization might be influenced and degraded by terminating the processes [12]. Another issue is to determine which processes are able to break the deadlock and solve the problem. Processes holding different resources dynamically and specifying the best set of processes that allocate enough resources to break the deadlock is a difficult task in such environment.

The objective of this paper is to find an optimal set of processes that can break the deadlock at minimum potential cost. Processes that can break deadlocks and whose termination will cause the minimum cost need to be specified and selected for termination. Thus, the problem of this paper can be represented in the following question:

- How can an optimal set of processes whose termination causes less impact on the overall system performance and can break a deadlock be selected?

Consequently, a goal programming (GP) model that seeks an optimal set of processes that can break the deadlock at minimum cost is proposed. Indeed, several selection criteria related to system performance and breaking deadlocks are proposed, and a GP model responsible for selecting processes that best achieve such criteria to be terminated is constructed. The remainder of this paper is organized as follows: the related works are reviewed in section 2. Section 3 presents the requirements size and a detailed design of the proposed solution. Evaluation result and discussion is introduced in section 4. Finally, some conclusions are given in section 5.

2. RELATED WORKS

Several varied strategies are used for recovery from a deadlock [11] [13] [14]. Process termination, especially partial termination, is one common solution that is used for deadlock recovery [3,15]. In partial termination, one or more of the deadlocked processes are aborted and its resources are released until the deadlock is broken [16]. Nevertheless, the primary issue is to find the best processes whose termination will result in minimum cost to be terminated [7-9,17].

Many researchers have found out that the primary step of deadlock recovery is to select a process and then to abort it [18-20]. Different attributes of processes or resources have been adopted when selecting a process for termination. The simplest method is to abort any deadlocked process [21-22], or randomly terminate any process until a deadlock is broken [14,23]. The authors in [24] discuss minimum cost recovery for multiprogramming by selecting processes based on the number of held and requested resources; in addition to a number of requests that a process can satisfies. The cost represent the computation complexity in terms of time.

Another method proposed in [25] selects processes based on a dynamic process priority that was assigned based on several attributes such as; process age, process history, process code size, process priority, resource utilized, number of in/out-degree edges, and cycle participation. The author in [26] proposes the use of heuristics for choosing the right process such as the last process that has been loaded, or the process with the least remaining processing time. Many other criteria are used to guide process selection for termination such as process code size [27], number of deadlock cycles that the process involved [15,28], number of submitted operations [29,30], number of holding resources [9,28], number of terminated processes [31-32], process age [32-33] and process priority [8-9,20,34-35].

However, the cost of aborting processes has been mainly measured in terms of either time complexity [9,15,30,36], or message complexity [6,8,37]. Process termination may lead to a significant degradation in system performance (e.g.
throughput and resource utilization) rather than time and message complexity. The investigation or measurement of potential cost of process termination as an impact on system performance is generally missing in the literature. As pointed out in [14], however, cost is a general term and several factors may affect the decision of which deadlocked process should be terminated. Therefore, measuring and minimizing the cost of process termination in terms of system performance is required and adopted in this paper.

In fact, some process selection methods aimed to improve a specific performance aspect instead of maintaining overall system performance. In this sense, one method can be ideal in one aspect of system performance, but not in other aspects. Successful recovery from a deadlock is the one that maintains the overall system performance encompassing several desirable performance aspects simultaneously. Furthermore, such methods do not actually model a recovery strategy from detected deadlocks. Several desirable performance parameters that could improve the overall system performance will be considered in this paper when measuring and minimizing cost of process termination.

Moreover, little attention has been given to recovery from deadlocks in systems that have multiple instances for each resource type in the literature. Deadlock detection and recovery in systems that have single instance of each resource type can be simply performed using a wait-for-graphs [13,15] which require less operations and overhead. In contrast, deadlock detection and recovery in systems that have multiple instances for each resource type is more challenging and the cost might be higher. Therefore, the proposed solution is intended to deal with resources that has multiple instances.

Consequently, previous solutions of recovery from deadlocks depend on a deadlock detection algorithm to determine whether the deadlock is resolved, and they are not concerned with the ability to break deadlock in their selection strategy. Such a procedure requires frequent execution of deadlock detection algorithm for each occurrence of a deadlock situation. Consequently, valuable system resources may be wasted as a result of frequent execution of deadlock detection algorithm [13]; which increase the overhead and cost. However, the proposed solution do not require a detection algorithm to insure that a deadlock situation is resolved. Alternatively, the GP selected the processes according to its ability to break the deadlock. This can reduce the overhead of detection that might be required after each process selection and improve the performance.

3. THE PROPOSED SOLUTION

The proposed solution for minimum cost deadlock recovery is to determine which deadlocked processes should be terminated based on its ability to break the deadlock, and based on the potential cost that might be resulted from aborting such processes. Simultaneous achievement of such constraints is required when performing the selection. The following subsections present the details of the solution.

3.1 Breaking a Deadlock

A deadlock occurs when there is not enough available resources that can be used by waiting processes. Each deadlocked process waits for resources that are occupied by other waiting processes. The goal of process termination is to release a number of resources that can satisfy the requests of one or more other waiting processes and leads to break the deadlock. Thus, the optimal set of processes holding resources that are enough to satisfy the requests of some deadlocked processes need to be specified and selected for termination. However, the proposed GP will be formulated to find a best set of processes that hold enough number of resources to break a deadlock to be terminated. This formulation guarantees recovery from the deadlock without need to use of deadlock detection algorithm.

3.2 Cost of Recovery from a Deadlock

Process termination may lead to interrupt resource usage, and to waste performed work and tasks for the terminated processes which eventually affect the overall system performance, especially throughput and resource utilization [11,14,25]. Thus, several cost factors - as pointed out in [14] - that might contribute to maintain the system performance will be used in the proposed solution to measure and minimize the cost of aborting processes. These factors will be modeled and formulated in this paper to be used as selection criteria that guide to a minimum cost deadlock recovery. The cost factors are presented in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Cost Factors (selection criteria) for Process Selection [14]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What is the priority of the process</td>
</tr>
<tr>
<td>2. How long time the process has computed</td>
</tr>
<tr>
<td>3. How long time the process still requires in order to complete</td>
</tr>
<tr>
<td>4. How many resources the process has used</td>
</tr>
<tr>
<td>5. How many resources left for the process to complete</td>
</tr>
</tbody>
</table>

Indeed, each factor can contribute in minimizing the cost by maintaining a specific performance parameter for the system. For example, reducing the number of wasted resources is desired and needs to be attained when recovering from a deadlock; thus the number of resources that the process has used needs to be considered as a cost factor in order to be minimized. Using the previous cost factors can contribute mainly in maintaining throughput, resource utilization and fairness among process, which are important and desired attributes in resource sharing and process cooperating system environments [1, 12]. Such performance attributes might be significantly degraded when recovering from deadlocks by process termination.

Fairness among processes can be achieved by selecting processes according to a predetermined priority rather than any other arbitrary reference. A system that achieves fairness among processes is better in terms of cost. Throughput can be maintained by either increasing the completed work or tasks in a given time or increasing the number of processes completed per unit time [1]. Indeed, processes that require more execution time and higher number of resources in order to complete are assumed to consume more resources and finish the execution much later, and terminating such processes can let smaller processes to finish in a given time which attain higher throughput. Furthermore, processes that have computed longer time are supposed to complete more work or tasks, thus keeping such processes and terminating processes that have computed less time can help in maintaining throughput. Moreover, higher performed work and completed tasks can indicate higher resource utilization. Thus, terminating processes that hold and use less number of resources reduce the number of wasted resources that are already being used, and keep more resources executing and completing their designated tasks rather than interrupting their execution which indicates better resource utilization. A system that achieves higher throughput and resources utilization is better at a lower cost.
3.3 The Goal Programming Model

The objective of this paper is to perform deadlock recovery at minimum cost by process termination. Thus, cost factors in addition to the ability of breaking the deadlock need to be used as selection criteria, and the selection of processes for termination will be performed up on these criteria. However, the challenge lies in finding a solution that best achieves such criteria concurrently; processes that are optimal in minimizing cost may not guarantee to break the deadlock and vice versa. Furthermore, a set of processes that is optimal in one cost factor as a selection criterion may not be optimal in others. The best solution in such case is always a compromise, and all metrics of interest must be taken into account concurrently.

Accordingly, a GP model responsible for finding an optimal set of processes that best satisfies such multiple selection criteria is proposed. GP is a multi-criteria satisfying methodology that seeks a solution that best fits or satisfies the desired set of criteria in a multi-criteria decision making problems [38].

In fact, the main objective function to be formulated in the GP model is to minimize the cost that has several factors. Consequently, percentage of execution time \(T\), percentage of resource allocation \(R\), and priority violation \(P\) are proposed as input variables for the GP model to help measure and minimize the cost that represents the proposed factors. \(T\), \(R\), and \(P\) variables can be calculated for each deadlocked processes based on available information that can be obtained dynamically from the underlying operating system at a time of performing deadlock detection and recovery. Through such variables, the proposed GP can find processes that best satisfy the adopted cost factors and minimize the cost. Table 2 shows the description of the proposed formulation variables and its relation to the cost factors.

<table>
<thead>
<tr>
<th>V</th>
<th>CF</th>
<th>Parameter</th>
<th>Description</th>
<th>GP selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_i)</td>
<td>1</td>
<td>Priority</td>
<td>Priority of the process.</td>
<td>Lower value</td>
</tr>
<tr>
<td>(T)</td>
<td>2</td>
<td>Time computed</td>
<td>Amount of time that the process has computed.</td>
<td>Lower value</td>
</tr>
<tr>
<td>3</td>
<td>Time required</td>
<td>Amount of remaining time that required for the process in order to complete.</td>
<td>Higher value</td>
<td></td>
</tr>
<tr>
<td>(R)</td>
<td>4</td>
<td>Resource allocation</td>
<td>Number of resources of each type currently allocated to each process.</td>
<td>Lower value</td>
</tr>
<tr>
<td>5</td>
<td>Resource required</td>
<td>Number of resources that required for each process in order to complete.</td>
<td>Higher value</td>
<td></td>
</tr>
</tbody>
</table>

\(T\), \(R\), and \(P\) variables can be calculated for a process \(i\) using the following equations.

\[
R_i = \frac{\text{Resource allocation }}{\text{Resource allocation } + \text{Resource required }} \quad (1)
\]

\[
T_i = \frac{\text{Time computed }}{\text{Time computed } + \text{Time require } d_i} \quad (2)
\]

\[
P_{hi} = \text{priority } i - \text{min}(\text{priority}) \quad (3)
\]

Where Min (priority) is a function that returns smallest value of priority for all deadlocked processes that exist in the system at a time of performing deadlock recovery.

As a result, minimizing the cost of aborting processes as a main objective can be divided into the following sub-goals as: (i) minimize the percentage of wasted execution time \(T\); (ii) Minimize percentage of wasted resources \(R\); (iii) Minimize priority violation \(P\).

Additionally, in order to select a set of processes that able to break the deadlock, set of related constraints are developed and used in the model. Finally, the following GP model is constructed as presented in Figure 1:

\[
\begin{align*}
1 & \text{ Min Cost;} \\
2 & \text{ Subject to:} \\
3 & \text{ Cost } = d_R^+ + d_R^- + d_T^+ + d_T^- + d_P^+ + d_P^-; \\
4 & (\sum_{i=1}^{n} R_i * P_i) - d_R^- + d_R^+ = 0; \\
5 & (\sum_{i=1}^{n} T_i * P_i) - d_T^- + d_T^+ = 0; \\
6 & (\sum_{i=1}^{n} P_i) - d_P^- + d_P^+ = 0; \\
7 & \text{ For } k=0, \ldots, n-1 \\
8 & \text{ For } j=1, \ldots, r \\
9 & (\sum_{i=0}^{r} \text{request}_{ij} * T_i P_i) \leq \text{available}_{ij} + (\sum_{i=0}^{r} \text{hold}_{ij} * T_i P_i) + T_{m=0,k=0} (\sum_{i=0}^{r} \text{hold}_{ij} * T_i P_i); \\
10 & \text{ For } i=0, \ldots, n-1 \\
& (\sum_{j=0}^{n} T_j P_i) + P_i = 1
\end{align*}
\]

Fig. 1 the proposed GP model for recovery from deadlocks

Where \(n\) is the number of deadlocked processes, \(r\) is the number of resource types in the system, \text{request}_{ij} is the number of requested resources of type \(j\) by a process \(i\), \text{hold}_{ij} is the number of allocated resources of type \(j\) by a process \(i\), \(P_i\) and \(T_i P_i\) are binary decision variables (0/1 variables), and \(d_R^+, d_R^-, d_T^+, d_T^-, d_P^+, d_P^-\) are deviational variables. The expected output of the model is a vector of 0/1 values corresponds to each process decision variable \(P\). 0 means that the correspondent process is not selected, where 1 means that the associated process is suitable for termination. Processes with 1 value are the optimal processes whose termination together as a set can break the deadlock at minimum cost.

4. EVALUATION RESULTS AND DISCUSSION

In order to measure the ability of the proposed GP model to perform deadlock recovery with the least possible cost, and to compare its performance with related methods, a simulation system was constructed and several tests were run. The simulation was executed to generate varied deadlock situations under the following assumptions:

- The number of allocated resources, the number of requested resources, the amount of computed time, the amount of required time, and the priority for each process were selected randomly.
- The simulation selected varied numbers of resource types \(r\) with random number of instances in each type.
- Each deadlocked process requests a number of resources less than or equal to the number the system has.
- All system resources are in use
- The number of requested resources were selected under the constraint \(\text{request}_{ij} < \text{hold}_{ij}\) less than or equal to all resources of that type in system.

However, deadlock situations were generated randomly under three different values of resource types, where \(r = 1, 4, \text{ and } 8\).
The number of requested resources by each process were selected randomly between 0 and MAXRequest. MAXRequest can reflect the relation between the number of allocated and requested resources to the number of all resources in the system. MAXRequest = factor*(request-hold). The factor reflects the degree of how costly the recovery from a deadlock [24]. The selected factor values were 0.1, 0.5, and 1.0. For each of the 9 combinations of r and factor values, priorities were generated randomly between 0 and MAXPriority where MAXPriority= 2, 10, and 20, respectively. This makes up three variations of priorities. For each of the 27 combinations of resource types, factors, and MAXPriority values, 250 deadlock situations were generated resulted in a total of 6750 deadlock situations.

In sum, all scenarios have been classified, summarized, and tested as depicted in Table 3 into three sets: Set (A) is to measure and compare the performance of the related methods in handling deadlocks with different number of resource types, factors, and MAXPriority values, set (B) is to measure and compare the performance of the related methods in handling deadlocks with different values of deadlock factors, and set (C) is to measure and compare the performance of the related methods in handling deadlocks with different variations of priorities.

Table 3. Simulation Scenarios

<table>
<thead>
<tr>
<th>No</th>
<th>Resource type</th>
<th>Factor</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test Value</td>
<td>Test Value</td>
<td>Test Value</td>
</tr>
<tr>
<td>1</td>
<td>(A)h 1</td>
<td>(B)h01 0.1</td>
<td>(C)h02 0-2</td>
</tr>
<tr>
<td>2</td>
<td>(A)h 4</td>
<td>0.1</td>
<td>0-2</td>
</tr>
<tr>
<td>3</td>
<td>(A)h 8</td>
<td>0.1</td>
<td>0-2</td>
</tr>
<tr>
<td>4</td>
<td>(A)h 1</td>
<td>(B)h05 0.5</td>
<td>0-2</td>
</tr>
<tr>
<td>5</td>
<td>(A)h 4</td>
<td>0.5</td>
<td>0-2</td>
</tr>
<tr>
<td>6</td>
<td>(A)h 8</td>
<td>0.5</td>
<td>0-2</td>
</tr>
<tr>
<td>7</td>
<td>(A)h 1</td>
<td>(B)h10 1</td>
<td>0-2</td>
</tr>
<tr>
<td>8</td>
<td>(A)h 4</td>
<td>1</td>
<td>0-2</td>
</tr>
<tr>
<td>9</td>
<td>(A)h 8</td>
<td>1</td>
<td>0-2</td>
</tr>
<tr>
<td>10</td>
<td>(A)h 1</td>
<td>(B)h01 0.1</td>
<td>(C)h10 0-10</td>
</tr>
<tr>
<td>11</td>
<td>(A)h 4</td>
<td>0.1</td>
<td>0-10</td>
</tr>
<tr>
<td>12</td>
<td>(A)h 8</td>
<td>0.1</td>
<td>0-10</td>
</tr>
<tr>
<td>13</td>
<td>(A)h 1</td>
<td>(B)h05 0.5</td>
<td>0-10</td>
</tr>
<tr>
<td>14</td>
<td>(A)h 4</td>
<td>0.5</td>
<td>0-10</td>
</tr>
<tr>
<td>15</td>
<td>(A)h 8</td>
<td>0.5</td>
<td>0-10</td>
</tr>
<tr>
<td>16</td>
<td>(A)h 1</td>
<td>(B)h10 1</td>
<td>0-10</td>
</tr>
<tr>
<td>17</td>
<td>(A)h 4</td>
<td>1</td>
<td>0-10</td>
</tr>
<tr>
<td>18</td>
<td>(A)h 8</td>
<td>1</td>
<td>0-10</td>
</tr>
<tr>
<td>19</td>
<td>(A)h 1</td>
<td>(B)h01 0.1</td>
<td>(C)h20 0-20</td>
</tr>
<tr>
<td>20</td>
<td>(A)h 4</td>
<td>0.1</td>
<td>0-20</td>
</tr>
<tr>
<td>21</td>
<td>(A)h 8</td>
<td>0.1</td>
<td>0-20</td>
</tr>
<tr>
<td>22</td>
<td>(A)h 1</td>
<td>(B)h05 0.5</td>
<td>0-20</td>
</tr>
</tbody>
</table>

Consequently, for each deadlock situation generated, the deadlock was recovered by terminating processes that selected using five methods namely; least required resources (LRR), least required processing time (LRPT), first detected first aborted process (order of detection), priority, and the proposed GP model. Each method selects a set of processes to be aborted and recover from each deadlock situation, and the cost was calculated based on the selected processes.

Figure 2 shows the performance of the GP model compared to the average performance of all other related methods under different system parameters. Firstly, the results showed that performing deadlock recovery becomes more difficult with the increase of the values of parameters. That is, the expected cost of deadlock recovery increases with environmental changes. Secondly, the average performance of the GP model is significantly better than average performance of the four related methods and it can scale to varying environmental changes at a minimum cost.

More specifically, as depicted in Figure 3 (a-d), the proposed GP method performs better than the other methods and recover all deadlock situations at minimum cost under all tests. This could be explained as the GP selects processes based on the potential cost that might be resulted and based on its ability to break deadlock simultaneously. Furthermore, GP seeks processes that best achieve all cost factors which have better result for the overall system performance. Comparatively, methods such as order selection do not take into account any selection criteria which significantly increased the recovery cost. Other methods that are based on specific criteria such as the number of required resources, remaining processing time, or priority consider only one cost factor which partially reduces the cost.
Figure 4 summarizes the overall cost achieved by all methods for all generated deadlock situations. The cost of aborting processes resulted when using the GP method was the lower among the other methods with varying degrees. The GP outperforms the LRR with 86% efficiency, LRPT with 87% efficiency, order with 60% efficiency, and priority with 24% efficiency. Furthermore, the detection algorithm was not required when performing recovery by the proposed GP model which significantly reduces the recovery cost in terms of overhead. In contrast, the detection algorithm was frequently executed for each deadlock situation when it was recovered by each other method.

**5. CONCLUSION**

Deadlock recovery at minimum cost has been addressed in this paper. The proposed solution was to find and terminate set of processes that best reduce the potential termination cost. Performance attributes such as throughput, resource utilization and fairness are used as selection criteria in addition to the ability to break the deadlock to help in choosing processes for termination. Accordingly, a linear goal programming model was proposed to find processes that best satisfy such multiple selection criteria. Measuring and minimizing recovery cost based on the proposed criteria reduced the negative impact of aborting processes on the overall system performance.

The performance of the proposed GP has been measured and compared to some related methods that might be used to select a process for termination. The result showed that the GP selected the best set of processes for termination that were able to break the deadlock and conducted minimum potential cost. The GP outperforms the other selection methods with 64% average overall efficiency. However, as a future development of the new model, the proposed selection criteria can be dynamically weighed according to their level of achievement in the system; and use these weights in the proposed GP model. This might best maintain such desirable performance criteria dynamically.

**6. REFERENCES**


[27] Weikum, G., & Vossen, G. Transactional information systems. ACM, 2002


