A Prescriptive Logic Model for Software Application Root Cause Analysis

Hoo Meng Wong, Sagaya Sabestinal Amalathas, and Tatana Zitkova

Abstract—Software application stops functioning in the production environment can be due to the error either within the software application layer or any other factor outside the software application layer. With the complexity of an enterprise level of software application involving multiple tiers, it has log file at each tier. The root cause analysis activity will easily prolong as more than one log file required for the root cause analysis activity. This would increase the total time taken on restoring the software application service back to the users. To identify the root cause of software application error in a more accurate manner, and shorten the duration of root cause analysis activity conducting on software application error. A Prescriptive Analytical Logic Model (PAL) incorporates with Analytic Hierarchy Process (AHP) is proposed. The logic model contributes a new knowledge in the area of log file analysis to shorten the total time spent on root cause analysis activity. At the same time, it contributes the knowledge in AHP to close the knowledge gap.

Index Terms—Algorithm, Analytic Hierarchy Process, Application Log File Analysis, Application Log File Debugging.

I. INTRODUCTION

Business companies adopt Information Technology (IT) as a strategic enabler to sustain their business operations. When the software application becomes crucial for processing the business transactions, it has lower toleration for downtime expected by the company management. This was clearly supported by Labels: Data Center, Downtime, www.evolven.com (2014), and indeed software application downtime can cause the business operation ceased. Business company has option that it can run an IT production support team to provide support service to whichever business-as-usual (BAU) system running in the organization. Another option is that it can engage service provider to provide the same IT support service. Regardless whichever option is chosen, the time spent on conducting root cause analysis on software application error is crucial. Without accurately identifying the valid error, it creates impact to the service restoration to the software application.

As of the fact, identifying the root cause of the software application error is crucial before the resolution is decided and deployed into production environment. There are several required actions in the root cause analysis activity. These actions are:

i. Collecting related information from different log files.
ii. Selecting the related log events based on the time event when software application error occurred.

Most of the time, collecting input information for root cause analysis is time consuming. This statement is supported by Management Logic (2012) stated that "The most time consuming aspect of Root Cause Analysis (RCA). Practitioners must gather all the evidence to fully understand the incident or failure.". On the other hand, Horvath, Kristof (2015) had also pointed out that "While the analysis itself can be time-consuming, the chance to mitigate or eliminate the root causes of several recurring problems / problem patterns is definitely worth the effort.". Hence, it is crucial to look for an efficient method to reduce the prolonging time at the root cause analysis activity.

Generally, each software application has its built-in event logging ability. The purpose of this event logging records the information of what activity is carried out or even what incident is occurred at that exact time. This information includes appropriate debugging information, and later the same information can be analyzed for software application root cause analysis purpose. The concern raised to the required information logging is that how much logging information is accepted as sufficient for software application root cause analysis. In addition, what is the appropriate category for the logging event such as information, error, debug, and fatal should be fetched as the input information to the root cause analysis activity. In the situation that if the extensive event logging level is enabled, this can lead to excessive logging information generated. With that, there are two issues raised.

i. The first issue is that, the performance of software application is reduced by comparing with before and after extensive event logging option is enabled.

ii. The second issue is that, the manual analysis activity is becoming much more difficult and even tedious to identify the root cause of the software application error.

Therefore, in the software application development process, it is a great concern on how much detail event logging should be logged into the log file. At the same time, the event logging must mitigate the performance impact created to the software application. These mentioned concerns had also been highlighted by Loggly (2017) and Panda, A (2011). As per the following figure 1.0, Operating System communicates between software application and assigned resources (such as CPU, Memory, Network, and Hard Disk) on the virtual machine. Software application has to interact with Operating System to obtain allocated server resources to handle software application processing. This is because software application has high dependency on server resources to carry out its execution. Without the server resources, software application cannot execute itself at the software application layer. Furthermore, software application requires to communicate to its database server.

Published on October 29, 2019.

DOI: http://dx.doi.org/10.24018/ejece.2019.3.5.133
their users on using all or partial of the software applications and re-investment budget on spending the capital amount to replace and at the same time to allow the companies to save the

time for the software application utilized
day software application error challenge in order to make a right or wrong judgment on a software application error.

(iii) The root cause analysis activity is conducted by human; it is very subjective to the person would make a right or wrong judgment on a software application error.

(iv) When software application support team becomes costly, it leads to high expectation from company’s management to have the lowest software application down-time. This has caused the job of software application support team to become extremely stressful.

The consequences of the reasons:
(v) The business users face difficulty to continue their daily tasks as the software application is unavailable or malfunction.

(vi) High expectation on software application support team by the company management to resume software application service when software application service or certain functions are unavailable.

IV. STATEMENT OF PROBLEM

The duration of root cause analysis on software application error carries crucial impact to the service restoration.

V. RESEARCH OBJECTIVE

To mitigate prolonging on conducting root cause analysis activity.

VI. LITERATURE REVIEW

In the past research, Stewart, D (2012) is focusing on debugging real-time software application error using logic analyzer debug macros, whereby Eick, S, Nelson, M, and Schmidt, J (1994) are focusing on presenting the error logs in a readable manner. Moreover, Wendy, P and Dolores, R (1993) suggested to focus on error detection in software application at the time of software development and maintenance. However, Salfner, F and Tschirpke, S (2015) are focusing on analyzing error logs by applying the proposed algorithms in order to predict future failure. Their software application error analysis approaches focus on software application error log obtained from software application database. Some literature suggested that the software application error analysis would be better if it is built in during the software development process. This approach is still within the same application development boundary without factoring in any other areas of concerns. It can cause the software application failure. In another way of explanation, whenever hardware CPU and memory utilization is running high, or even storage disk space is running low, software application logging may not be accurate anymore. Hence, the root cause analysis would not be accurate to identify the real issue to understand the main reason to cause the software application failure. Murinová, J (2015) had attempted to integrate multiple log files from various software monitoring tool and network devices for better root cause analysis on Web application error. However, there is no proposed model stated in the research. Even until Landauer et al (2018), they introduced an unsupervised cluster evolution approach that is a self-learning algorithm that detects anomalies in terms of conducting log file analysis. However, this approach is under machine learning rather than AHP. From a different point of view, this approach is good because it can be adopted into the proposed model to detect the software application error. Hence, this is a great potential to propose a logic model to research a new approach towards developing an algorithm for software application error analysis. At the same time to contribute new knowledge in the area of software application root cause analysis using AHP. By comparing the above secondary data with the proposed logic model, it can be noticed that the focus boundary on software application analysis. The technique to identify the root cause of software application is different. They focus on software application boundary whereby the proposed model focuses horizontally on all possible boundaries. In addition, due to the focus boundary is different, it leads to the technique to identify the root cause of software application also different.

Although there are many market products available for review, as for the literature review, we choose the most popular three market products for features comparison. The product information in the following tables are referred by Solarwinds (2018), Logstash (2018) and AppDynamics (2018).

<table>
<thead>
<tr>
<th>TABLE I: MARKET PRODUCT COMPARISON FOR MONITORING AND ANALYZING LOGS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Product</strong></td>
</tr>
<tr>
<td>Features</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

On the other hand, Kanojia (2018) proposed a product which is named Micro Focus Operations Bridge, this product will sit on top of all the existing monitoring tools, and to consolidate events and correlate them in real time. This product introduces an Analytic Engine which can perform predictive analytic based on the data that has been collected over a period (historical data from log files and/or events) and claims that it can predict issues before the issue occurs.

From the above table, it is easily to understand from the product feature comparison in term of the common features. The common features are such as collecting all possible logs into a central location, and for each type of collected log, it gets massaged and presented into a graphical view under a widget. After all, by putting all the widgets together into a so-called dashboard, it creates a bird-view to oversee all software application status in a graphical visualization for better viewing. In addition, the market product allows users to configure predefined resolution action for the standard software application error. However, the decision making on identifying the valid software and predefined resolution steps are still on human.

VII. RESEARCH PROPOSED

With a server (regardless it is physical or virtual) box, you have multiple layers. Most common logs that are required is:

(a) System monitoring log for server resources such as CPU, Memory and Hard Disk usage.

(b) Network monitoring log for the server such as network communication within the Local Area Network (LAN).

(c) Operating System event log for server.
The illustration of multiple layers for a server can be referred to Figure 2.0. On the other hand, the software application especially for those that are rated at enterprise level category, it involves multiple tiers such as Web Client tier, Web Container tier, Application Container tier, and Database tier. Beside the Wen Client tier, each tier has its own log file. Certain tiers are even having multiple log files based on the software application design. The proposed research is to develop an algorithm towards to the PAL for analyzing software application error. The analysis is based on the logs retrieved from various software applications by referring to the following figure. With this proposed logic model, it is aiming to resolve the statement of problem, which is prolonging the time consumption on identifying the root cause of the software application error would increase the total time taken on restoring the software application service to the users.

A. Proposed Research Scope

The proposed scope of this research is to define the algorithm. The algorithm consists of simple and complex analysis inside the PAL for software application error analysis. Therefore, by having the proposed logic model in the production environment, the PAL is required to react to software application error when the error is detected in the software application log file. With this PAL, it is also required to retrieve other related log files through various software applications shown as Figure 2.0. The proposed algorithm mainly consists of two analysis areas, which are simple and complex analysis to form a prescriptive analytical logic.

B. Proposed Simple Analysis

For simple analysis area, it is required to build the predefined logic to handle the common software application error, whereby this model is to guide the system builder on answering a set of predefined questions on common software application errors and carry out the predefined activities only react to these common software application error, for example, restarting the software application process if it is stopped.

C. Proposed Complex Analysis

For complex analysis area, it is required to build a logic which collects necessary log events as data from the involved software applications. The collected data will serve as input information at the initial stage. With the collected data, this model will base on the analysis of root cause for the system behavior. By combined with logical model templates, the automated analysis activities will be triggered and finally generate the analysis outcome along with the suggested resolution steps and action to the IT support team. This complex analysis would have three different modes which are “manual”, “semi-auto” and “fully-auto” offered to the IT support team. As for the complex analysis area, by predicting the software application behavior, it performs the suggested steps and carry out the action against the software application error based on the analysis. This will prevent future application failure based on the permission given to the offered mode by the IT support team.

By focusing into the re-occur software application errors, these errors occur in a specific pattern or feature, and the solution is often straightforward (can be applied after validating the specific pattern or feature) to resolve the incidents. The human involvement on this type of incidents would require less analysis but more on validating activities. Hence if the validating activities can be predefined into a checklist, the PAL is able to pick up the ultimate predefined solution and react to the incident automatically. This can be achieved by the combination of the answers (yield from the validating activities in the checklist). This would be the preferred method in the PAL that handles the common software application incidents. We call this logic as simple analysis. The same simple analysis logic can be applied to manage Server (a physical or virtual box running a vendor Operating System) or even Networking devices (such as switch or router) if they have incidents occur in the specific pattern or feature.

A certain percentage of errors that are occurring will have no uniform pattern or feature, for this type of software application errors. The percentage of human involvement is high. This is because the person who handles the incident requires to obtain the software application log files and to search any similar error logged in the past. We call these files and records as input information. With the input information obtained, the person conducts the analysis activities before the person can identify the software application error root cause. Only the preferred resolution steps is agreed then it is applied to resolve the software application error.

For the first time occurring software application error. If both yielding input information activities and analysis activities can be automated. Base on the outcome of the analysis activities, human expects to see a list down of each possible root cause along with the proposed resolution steps in a complete list. Then, the decision is on the person to choose which is the the preferred option. If the person chooses to proceed with the suggested resolution steps, then the person will receive the final question. The question is
expecting the response from the person, whether agrees to let the automated activities execute the same suggested resolution steps automatically in the future if the same incident occurs again. Of course, this logic has the ability to handle unpredictable software application errors by performing simulated analysis activities comparing with human, we call this logic as complex analysis.

Whenever the complex analysis is triggered. It will pull the related logs based on specific time frame (duration) before and during the software application failure from various application logs. These logs are:

- Software Application logs is the beginning to trigger the root cause analysis,
- Configuration management logs is for understanding any recent applied software application patches or Operating System patches,
- Performing and capacity monitoring logs is for identifying any hardware resources running insufficient, and
- Production support ticketing tool logs is for cross-checking any related issue recently occurred under the predefined database scheme.

These above logs as input information will be utilized crucially for root cause analysis to resolve the software application issue. Indeed, the simple analysis can be existed independently at the initial stage. However, when the specific number of reoccur incidents hits. The complex analysis will be activated to perform the required analysis activities automatically. It will produce the complete analysis report and suggestion(s). Base on this suggested design, the complex analysis would have a loosely but it is fairly important relationship with the simple analysis. This is because the complex analysis needs to understand how many times the simple analysis has handled the same incidents in the past. This information is crucial to make a decision on suggesting the reasonable resolution steps to the human after the complex analysis produces the analysis report.

D. Proposed Algorithm and Methodology

The proposed algorithm under the PAL includes the following crucial activities:

<table>
<thead>
<tr>
<th>No.</th>
<th>PAL Proposed Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>With the granted software application access, use “time” field as the linkage to retrieve and relate all the necessary log data across multiple databases.</td>
</tr>
<tr>
<td>2</td>
<td>Identify whether the newly reported software application error is first-time occurrence or re-occurrence by cross checking the database which is associated to the PAL.</td>
</tr>
<tr>
<td>3</td>
<td>Identify possible log data and select the necessary log data for analysis under the defined software application error classification.</td>
</tr>
<tr>
<td>4*</td>
<td>Allocate weight to each possible software application error based on Analytic hierarchy process (AHP).</td>
</tr>
<tr>
<td>5*</td>
<td>Shortlist the software application error under the highest weight.</td>
</tr>
<tr>
<td>6</td>
<td>Analyze the selected log data for shortlisted software application errors and define possible resolution option.</td>
</tr>
<tr>
<td>7*</td>
<td>Allocate weight to each possible resolution option based on AHP.</td>
</tr>
<tr>
<td>8*</td>
<td>Shortlist the preferred resolution option under the highest weight.</td>
</tr>
<tr>
<td>9</td>
<td>Deploy the preferred resolution option to fix the software application error under the predefined condition.</td>
</tr>
<tr>
<td>10</td>
<td>Store the analysis result and resolution action into a database which is associated to the PAL, for future reference and knowledge base activities.</td>
</tr>
</tbody>
</table>

The methodology of the PAL is derived from the proposed algorithm. The figure is shown as follows:

![Fig. 3. The methodology of Prescriptive Analytical Logic Model.](image)

The PAL algorithm design consists of simple and complex analysis design, proposed analytic hierarchy process design and knowledge based database design. From the simple and complex analysis design, it further derives to top-down and bottom up design. All of these designs will form PAL algorithm design as the complete model design. The PAL consists of two configuration files that are the brain of the PAL to identify the known errors and the action of the preferred resolution.

VIII. ANALYTIC HIERARCHY PROCESS (AHP)

AHP was developed by Thomas L. Saaty stated in Wikipedia (2015). The proposed algorithm of the prescriptive analytical logic can adopt the decision making process of AHP to decide the valid software application. Then, followed by using the same decision making process to shortlist the best resolution. This is because by referring to R. W. Saaty (1987), the three principles which are the decomposition principle, the comparative judgments, and the synthesizing priorities. In addition, Vaidya and Kumar (2004) had provided the discussion on how to apply AHP under the analytic hierarchy process. This helps to understand how the proposed algorithm applies in the scenario after the valid software application error, and preferred resolution are identified during the root cause analysis activity.

Based on those past researches or similar area of researches in the area of software application log file analysis. The AHP was not introduced or applied for handling the activities such as shortlisting the valid software and shortlisting the preferred resolution. Since AHP is a decision making process as a tool, it can be utilized to weight the both the possible root causes and possible resolutions under a hierarchical structure. Then, narrow down to the final root cause, and decide the best resolution among the shortlisted resolutions. Ideally, the AHP will be utilized at the following two (highlighted) specific proposed activity under the PAL, which are at the row 4, 5, 7 and 8 in the following Table.

By referring to the proposed activities in table 2.0, that are activity number 4, 5, 7 and 8 (with the Note *) in the above table. These proposed activities in the algorithm play an important role of AHP process. The activity details are explained respectively as follows:

The activity number 4 helps to identify the possible software application errors and to filter out all the false alarms. On those possible software application errors, each error will be further identified by the its error characteristic.

DOI: http://dx.doi.org/10.24018/ejece.2019.3.5.133
and categorized into specific error category under a predefined software application error category list.

The activity number 5 will perform assigning the weight to each possible software application error based on the error category the impact level of each error within the category. With the assigned weight on each error, it would easily shortlist the highest priority of software application error as the crucial error to be fixed.

The activity number 7 handles the process of assigning weight on each possible resolution after the analysis activities has been conducted at section 6. This is because possible circumstance can be happened when two similar resolutions are selected, but we need to identify which is the most suitable resolution can be applied to resolve the software application error.

The activity number 8 helps to identify the final preferred resolution to be applied for the crucial error after evaluating the weight, and this action will isolate multiple resolution actions to be applied to the crucial error to be fixed.

With the PAL, it carries high potential on knowledge contribution as the new logic model has the algorithm to be able to identify the root cause of the software application error more accurately under the AHP processing approach, and deliver the outcome to shorten the analysis duration of conducting root cause analysis activity during the software application downtime.

IX. PROPOSED PREREQUISITES OF PAL

A. Multiple Tiers Environment

In this proposed research, it is required to run on a multiple-tier environment which involves Client tier, Web Container tier, Application Container tier, and Database tier. The proposed algorithm will be implemented into a software plug-in component sitting at the logic tier to integrate all the required log event as input data from various software applications and store in a separate location for information retrieval later. A new standard database schema may be required on this research, and it will be applied across all the required databases of the involved software applications. This will help on retrieving log events as data from various databases when the root cause analysis activity is triggered at the logic tier whenever software application error is detected in the main software application log file. Otherwise, by granting necessary permission to retrieve log events as data from the involved software applications should be sufficient.

B. Network Time Protocol (NTP) Server

Whenever a software application involves in a multiple-tier environment, it is good to standardize the server time across all the involved servers. The supporting reason is that the proposed PAL is required to collect log data from different software applications based on the specific given time as the primary key, and therefore the server time on all the involved servers are required to have the network time synchronization. As per Masterclock (2016), “Accurate time stamping is key to root-cause analysis, determining when problems occurred and finding correlations. If network devices are out of sync by a few milliseconds or, in extreme cases a few seconds, it can be very difficult for network administrators to determine the sequence of events.”.

C. Technologies

The proposed prototype of the software plug-in component will be coded using Java Programming as this programming language is platform independent. At the platform, it required to install Java Runtime Environment (JRE) and configure the Java home path and memory allocation for the JRE.

X. CONCLUSION

There are many software application error analysis tools available in the IT industry today. However, these tools mainly analyze based on the collected software application error logs which is reasonable if the hardware resources of the system are under a healthy stage. On the other hand, software application comes along with its event logging feature. This feature is mostly embedded into the software application and its design is based on the situation that software application has adequate server resources. In the scenario whenever server resources are running low, software application will face issue during its execution. Other issues can even come from network connectivity, no response from database server and etc that is beyond the software application layer. Hence, if the root cause analysis is solely depending on software application error log file alone, it is definitely insufficient and the result cannot be accurate. By improving the software root cause analysis technique into a more human-like intelligent, a proposed logic model is required. Therefore, the proposed logic model that incorporates with AHP is a great potential to fill up the knowledge gap in AHP.

ACKNOWLEDGMENT

I would like to thank to Tatana Zitkova, for providing encouragement on my proposed research and PhD studies.

REFERENCES


DOI: http://dx.doi.org/10.24018/ejece.2019.3.5.133
Hoo Meng Wong is a PhD student in Taylor’s University, Malaysia. He holds two bachelor degrees and a master degree. The bachelor degrees are Bachelor of Science in Computer Science which is awarded by La Trobe University, Melbourne, Australia. The another bachelor degree is Bachelor of Business (Hons) in Banking and Finance which is awarded by Wawasan Open University, Malaysia. For the master degree, it is Master of Science in Technology Management which is awarded by Staffordshire University, United Kingdom. Hoo Meng is new to publishing journal paper although he had started the PhD studies back in 2016. At that time he was with UNITAR International University, Malaysia. The reason for him transferred to Taylor’s University because his supervisor had decided to join Taylor’s University as one of the outstanding universities in Malaysia. Hoo Meng is currently conducting a research and preparing a thesis with the title Prescriptive Analytical Logic Model on Software Application Error Analysis. This topic of research is focusing on decision making performed on two activities, that are first activity is to identify the valid software application error from the root cause analysis activity, and second activity is to identify the preferred resolution based on the involvement of software development life cycle (SDLC) on the fix. Indeed, the research is to proposed a new algorithm for the Prescriptive Analytical Logic Model. The decision making process is incorporated with the Analytic Hierarchy Process (AHP).

**References**


