# Reliability Enhancement from http Log Files In Composite Web Services

**P.Maruthurkarasi (alias) Rohini, C.Jayaprakash, R.Balaji Ganesh**

***Abstract— A Hybrid Reliability Model based on log analyzer is designed to evaluate the Reliability of Composite Web Services. Based on Dependability, Atomic Web Services are composed with a Central Co-ordination Function (Broker). Real Time Server Log Files are fed as input to the system. Log Analyzer reads the log entries and separates the individual response of the server along with the time stamps. Base on the Frequency of the service response are classified and the Error Rate is calculated by the difference in the Uptime and Downtime Stamps. The Broker designs and decides the acceptance of the service based on Error Rate (MTBF, MTTF, MTTR) and Fault Tolerance. As, the Error Rate and Service Reliability are inversely proportional, the server with low error rate provides high reliability. Our Experimental Results with their groupings prove that the reliability can be evaluated using the Web Log File analysis.***

***Index Terms— Composite Web Services, Error log, HTTP Status Error, Reliability, Web error.***

## I. INTRODUCTION

The prevalence of the World Wide Web also spreads intended or unintended problems on an ever larger scale. The problems include various malicious viruses as well as unintended problems caused by communication breakdowns, hardware failures , and software defects. Identifying the root causes for these problems can help us understand their severity and scope. More importantly, such understandings help us derive effective means to deal with the problems and improve web reliability.

Web Service is a method of communication between two electronic devices over the World Wide Web. The W3C defines a “Web service” as “a software system designed to support interoperable machine-to-machine interaction over a network”[1][2]. It has an interface described in a machine-process able format (specifically Web Services Description Language, known by the acronym WSDL).

Web services are frequent accessed over a network, such as the Internet, and executed on a remote system hosting the request services, Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.

### *A. Quality of Web Service Reliability (QoWSR)*

The Quality of Service is the important factor in the Web Services, to increase the Quality of the Services Reliability of services has to be maintained in high[2]. The Reliability of the Service is depending on the Fault tolerance capacity of the service. Generally, Services are not available at any time for the service consumer or the services does not response to the server because of some error, at that time the service consumer could not get the Services for their need.

The Problem can be rectified with the Log File Analyzer.

### *B. Log File*

The log file is text file. Its records are identical in format. Each record in the log file represents a single HTTP request. A log file record contains important information about a request:

- Host name or IP address
- The Date and Time of the request
- The Requested file name
- The HTTP response status and size
- The Referring URL
- The Browser Information

A browser may fire multiple HTTP request to Web Server to display a single Web page. This is because the Web pages needs different files including HTML document, images and JavaScript files. All these files require HTTP requests and the success response of the server displays the content correctly.

### *C. Log File Analyzer*

The Log File Analyzer is a tool to analyze the time stamp of the Log file or Web server HTTP Response or Request Status[3]. On application of broker model, the analyzer provides the uptime, downtime, failure rate of the server. The broker model is a logical entity capable of extracting the information from the log file and calculates the reliability evaluation parameters. These parameters are used to calculate the Reliability Analysis [4]. Some of the reliability evaluation parameters used in this work are MTBF, MTTR, MTTF. Reliability of the Web Services are analyze through this parameters. If the Parameters have fewer values means that Service was declared as the Reliable services. Doubly Stochastic Process is used for the calculating this parameters.

The broker of the server analyze all the services, the service which have less MTBF, MTTF, MTTR value as provided to the service consumer [4][5]

The rest of the paper is organized as follows: Section 2 explains the Web Server Log Files. Section 3 portraits the process of Web Log Analysis. Section 4 shows the experimental results of Reliability Evaluation. Section 5 describes the Conclusion and Future work of this process.

## II. WEB SERVER LOG FILES

Growth of Web Server has become a vital part of the business model. Internet web servers must be reliable, as they are truly international 24*7*365 sales mechanism, but Internet Web Servers are become slowly in their performance

Manuscript received April, 2013.

**P.Maruthurkarasi (alias) Rohini**, PG Scholar, M.Tech., IT Department of Information Technology, K.L.N. College of Information Technology, Pottapalayam, Sivagangai.

**C.Jayaprakash**, Associate Professor, Department of Information Technology,K.L.N. College of Information Technology, Pottapalayam, Sivagangai

**R. Balaji Ganesh**, Assistant Professor, Department of Information Technology, K.L.N. College of Information Technology, Pottapalayam, Sivagangai.

because of heavy workload of the server[2][5]. To evaluate the reliability and availability of the web servers, we need to analyze the log files of corresponding web server. To analyze the Web server, we should monitor the log files of the web server.

### *A. Content in Web Server Log Files*

A server log is a log file automatically created and maintained by a server of activity performed by it. The W3C maintains a standard format for web server log files, but other proprietary formats exist. More Recent entries are typically appended to the end of the file. Information about the request, HTTP code, bytes served, user agent, and referrer are typically added. These data can be combined into a single file, or separated into distinct logs, such as an access log, error log, referrer log. There are two more log files like Transfer Log, Agent Log[2][3]. These two log files are standard log files. The Referrer log or Agent log are used to create the extending log file format.

The Log Files consist of the Information about web Server. Basically, Web Log Files are the history of Web Servers. The Starting and Ending time stamps of the Web Server, Server IP address, Client IP address, Port number, HTTP Status codes and further more information are present in the Web Log files. Common Log File format consists of the following information.

- **Client IP** usually the IP address, but can be resolved to DNS by the server (not recommended).
- **File** requested by client (including directory)
- **Method** used in request (GET, POST, etc.)
- **Return Code** tells about the service was succeed or not, If it is fails means the return code specifies reason for failure.
- **Bytes Sent** back to the client in the response,
- **Referring URL** specifies the link to the users request.
- **Browser String** tells that browsers used by the users.
- **Authenticated username** is mentioned in the Log File
- **Cookie** related to the transaction.
- **Bytes Received** by the server in the request.
- **Time Taken** by the server to process the request**.**

### *B. Common Log Format (CLF)*

The Common Log Format also known as the NCSA Common log format is a standardized text file format used by web servers when generating server log files. Because the format is standardized, the files may be analyzed by a variety of web analysis programs. Each line in a file stored in the Common Log Format has the following syntax:

```
127.0.0.1 user-identifier frank [10/Oct/2000:13:55:36 -0700] “GET /apache_pb.gif HTTP/1.0” 200 2326.
```

**Figure 1. A Sample Log Entry**

**127.0.0.1** is the IP address of the client (remote host) which made the request to the server. **User-identifier** is the RFC 1413 identity of the client. **[10/Oct/2000:13:55:36 -0700]** is the date, time, and time zone when the server finished processing the request, by default in string time format %d/%b/%y:%h:%s%z.“**GET/apache_pb.gif HTTP/1.0**” is the request line from the client. The method **GET**, ***/apache_pb.gif*** the resource requested, and **HTTP/1.0** the HTTP protocol. **200** is the HTTP status code returned to the client. 2xx is a successful response, 3xx a redirection, 4xx a client error, and 5xx a server error. **2326** is the size of the object returned to the client, measured in bytes.

### *C. Error Log*

The Error Log File contains information about the error occurred at the server. If any Error occurs in the server the Server reports the event to the log file. The log file stores all the Information about the Error as an Error Log.

```
2002-07-05 18:45:09 172.31.77.6 2094 172.31.77.6 80 HTTP/1.1 GET /qos/1kbfile.txt 503 – ConnLimit
```

**Figure 2. A Error Log Entry**

The first set of log entries consist of the date and time of the occurrence of error. The second set of log entries consist of the IP address and Port number of the client and server. The Third level of log entries consists of the Protocol version of the server and client. Next level said about the type of the request and protocol status of the server. The client makes the Error, so the severity of the Error is used to restrict that type of error.

### *D. Access Log*

The Access log contains detailed information about client and directory connections. A connection is a sequence of requests from the same client with the following structure:

- Connection record, which gives the connection index and the IP address of the client.
- Bind record.
- Bind result record.
- Sequence of operation request/operation result pairs of records.
- Unbind record.
- Closed record.

```
[21/Apr/2007:11:39:51 -0700] conn=11 fd=608 slot=608 connection from 207.1.153.51 to 192.18.122.139
```

**Figure 3. A Access Log Entry**

The First set of log entries consist of the Date and Time of the server. The second set of the log entries consist of the connection number, file descriptor and slot number. The next level of the log entries consist of the IP address of the client and server.

## III. WEB LOG ANALYSIS

### *A. Characteristics of Web Service Reliability*

Key to the satisfactory performance of the web is acceptable reliability. The reliability for web applications can be defined as the probability of failure -free web operation completions. Acceptable reliability can be achieved via prevention of web failures or reduction of chances for such failures [1]. We define web failures as the inability to obtain and deliver information, such as documents or computational results, requested by web users. This definition conforms to the standard definition of failures being the behavioral deviations from user expectations (IEEE, 1990). Based on this definition, we can consider the following failure sources in this process of obtaining and delivering information requested by web users[5]:

- **Host or network failures:** Host hardware or systems failures and network communication problems may lead to web failures. However, such failures are no different from regular system or network failures, which can be



International Journal of Engineering and Advanced Technology
IJEAT
www.ijeat.org
Exploring Innovation

analyzed by existing techniques. Therefore, these failure sources are not the focus of our study.

- **Browser failures:** These failures can be treated the same way as software product failures, thus existing techniques for software quality assurance and reliability analysis can be used to deal with such problems. Therefore, they are not the focus of our study either.
- **Server or content failures:** Web failure s can also be caused by the information source itself at the server side. We will primarily deal with this kind of web failures in this study.

### *B. Log File Dataset*

In this paper, we analyze the web logs from Internet Traffic Archive (ITA). The Internet Traffic Archive is a moderated repository to support widespread access to traces of Internet network traffic, sponsored by ACM SIGCOMM. This archive utilizes Apache Web Server, a popular choice among many web hosts, and shares many common characteristics of web sites and web servers.

The traces can be used to study network dynamics, usage characteristics, and growth patterns, as well as providing the grist for trace-driven simulations. The archive is also open to programs for reducing raw trace data to more manageable forms, for generating synthetic traces, and for analyzing traces. These two traces contain two months' worth of all HTTP requests to the NASA Kennedy Space Center WWW server in Florida was taken into account for web log analyzer.

Some pre -existing log analyzers were tried to analyze the access logs. However, these analyzers only provide very limited capability for error analysis. Therefore, an adapted log analyzer was implemented in java to count the number of errors, grouping of error patterns, and to classify the frequently navigated patterns therein.

### *C. Analysis of Time Stamps*

Log File is analyzed for finding the error rate and reliability of the web server. Time Stamps of the log file are used to deduce the error rate and classify the error based on the response code of the server. From the time stamps the uptime and downtime of the service request is gathered. The server response is in the form of HTTP response code. The HTTP status code revels whether the server' response is success or failure. The severity of the server is stored in the corresponding log entry. Grouping of HTTP response code makes the process of reliability calculation easier. The timestamps are further fed as input to calculate the various reliability parameters.

### *D. HTTP Server Status Code – A Review*

After, the service request is processed by the web server, the server in-turn returns the service response code. The Internet Assigned Numbers Authority (IANA) maintains the official registry of HTTP status codes.

- **1xx Informational** - HTTP_INFO- This class of status code indicates a provisional response, consisting only of the Status-Line and optional headers
- **2xx Success** - HTTP_SUCCESS- This class of status codes indicates the action requested by the client was received, understood, accepted and processed successfully.
- **3xx Redirection** - HTTP_REDIRECT - This class of status code indicates that further action needs to be taken by the user agent to fulfill the request.
- **4xx Client Error** - HTTP_CLIENT_ERROR- This class of status code is intended for cases in which the client seems to have erred
- ***5xx Server Error*** - ***HTTP_SERVER_ERROR-*** Response status codes beginning with the digit "5" indicate cases in which the server is aware that it has encountered an error or is otherwise incapable of performing the request

In this work, focus has been made to list and analyzes the number of errors that fall under the group of 5xx Server Error. By inferring the number of server errors with time stamps, we can able to obtain the error rate of the server with an option of sub grouping of the type of server error occurred. This information will be helpful to mathematically deduce the reliability of the web server.

This status classification is used to analyze the capability of the corresponding server in terms of different performance evaluation parameters like MTTF, MTBF. The Error arrival rate can be calculated by using the grouped HTTP status code. If the error arrival rate is high in the server then reliability rate of that server will be low. The failure or error arrival rate is inversely proportional to the reliability rate.

### *E. Reliability Evaluation*

The failure information, when used in connection with workload measurement, can be used to evaluate the website reliability in the software reliability model. In this paper, we use the doubly stochastic model to access the website current reliability. The formula used to find the Reliability is shown in equation (1)

$$R_i(x) = e^{-\lambda_i x} \qquad (1)$$

where,

Ri = Reliability of $i^{th}$ web server
x = Random variable presenting number of outages
e = exponential distribution
λ = error rate for given time period

A doubly stochastic model is a type of model that can arise in many contexts, but in particular in modeling time series and stochastic process. The stochastic process or sometime random process is a collection of random variables. This is often used to represent the evaluation of some random value or *web service system over time*.

The Reliability of a server can be measured by calculating the three parameters as given below

- MTBF(Mean Time Between Failure)
- MTTF(Mean Time To Failure)
- MTTR(Mean Time To Repair)

***Mean Time Between Failure*** is a reliability term used to provide the amount of failures per million hours for a service. The MTBF can be calculated using equation (2)

**MTBF = uptime/number of breakdowns (2)**

where, number of breakdowns defines the number of errors grouped by status codes.

***Mean Time To Failure*** is a basic measure of reliability for non-repairable systems. MTTF is a statistical value and is meant to be the mean over a long period of time and across large number of servers. The MTTF can be calculated using equation (3)

**MTTF = 1/Failure Rate (3)**

***Mean Time To Repair*** is the time needed to repair a failed

module. In an operational system, repair generally means replacing a failed hardware part. But in web services, MTTR could be viewed as mean time to redirect the failed request to a different server. The MTTR can be calculated using equation (4)

**MTTR = downtime / number of break downs (4)**

The uptime is the timestamp at which the client sends the service request to the server; downtime is the timestamp at which the server sends the response to the client. Failure rate is the average error log of the server; number of breakdowns is the number of failure work that occurs in the server during the given time period. All these details are inferred from the log file and fed as input to calculate the reliability of the server using the above equations.

## IV. EXPERIMENTAL RESULTS

### A. *Log Analyzer*

The Web server log files consist of the billion number of log entries as defined in the section III (B). The Log Entries consists of the IP Address, Date and Time of the request, Requested file name, HTTP response status and size, Referring URL, Browser Information. The content of a log file contains both access log and error log is shown in fig 4.

From the information the Arrival time of the request has been extracted and that input is used to calculate the reliability parameters. An adaptive log analyzer based on matching regular expression has been developed in java platform. This Log file analyzer groups the log entries based on the error status code. The output for the log analyzer which extracts the information from the log file is shown in figure 5.

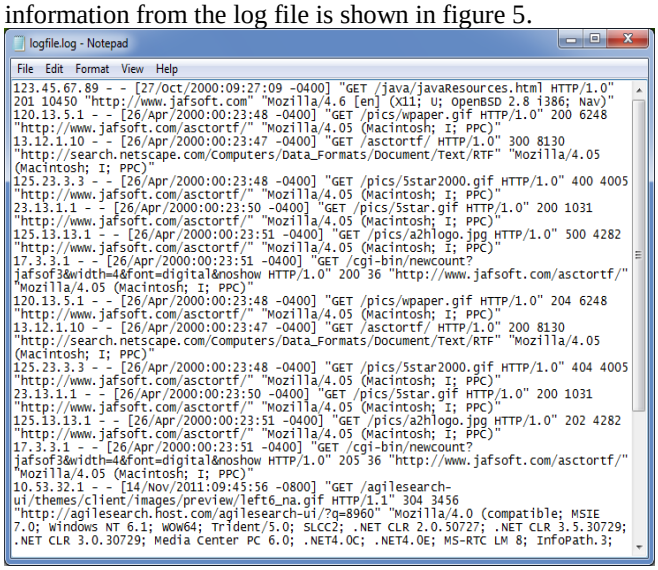


**Figure 4. Sample of Log Entries in Log File**

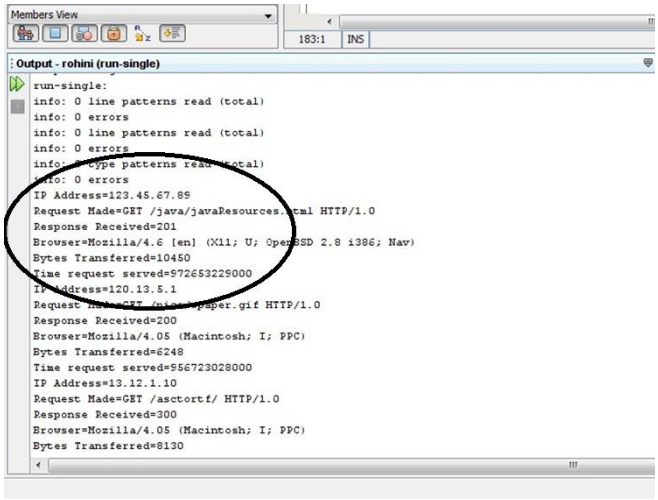


**Figure 5. Extraction of Information from Log File**

### B. *Reliability Analysis*

The reliability evaluation based on web log analysis was developed in the Java Platform and on a multiple server workspace. The service reliability of the servers was evaluated. For the purpose of evaluating the reliability, a workload table has been designed with service selection parameters. The results of reliability for different service request and its response from different servers are analyzes and the result sown in the Table I.

Table I.
Reliability Evaluation Table

| Service # | MTBF | MTTR | MTTF | Avail | Reliability |
|---|---|---|---|---|---|
| 1 | 0.286 | 0.571 | 0.12 | 0.33 | 1.234 |
| 2 | 0.6 | 1.4 | 1 | 0.3 | 0.0183 |
| 3 | 2.4 | 0.8 | 1 | 0.749 | 0.0497 |
| 4 | 6 | 8 | 0.125 | 0.428 | 1.266 |
| 5 | 0.834 | 5.67 | 1 | 0.128 | 1.234 |
| 6 | 0.75 | 0.438 | 0.34 | 0.631 | 3.059 |

Each response code has been grouped as shown in figure 6. In the figure the error status codes 3xx, 4xx, 5xx are grouped along with the number of occurrences. From the grouped response, the server error response was filtered (Response code 5xx). The uptime and downtime of the server error is given as input to calculate MTTF, MTTR, and MTBF. The average Error Arrival time of server error code has been consider as the MTTF and Arrival time of server error has been consider as the MTTR and difference between first and next server error as consider as the MTBF.

The time bound log entries (specifically server error entries) along with the workload information of the server can be used to evaluate the reliability of the server in terms of the three evaluation parameters. This evaluation actually reflects the dynamic status of the server.

The graph in figure 7 depicts the performance and reliability of the server. The Service Request Performance is directly proportional to the reliability and availability values.



International Journal of Engineering and Advanced Technology
IJEAT
www.ijeat.org
Exploring Innovation

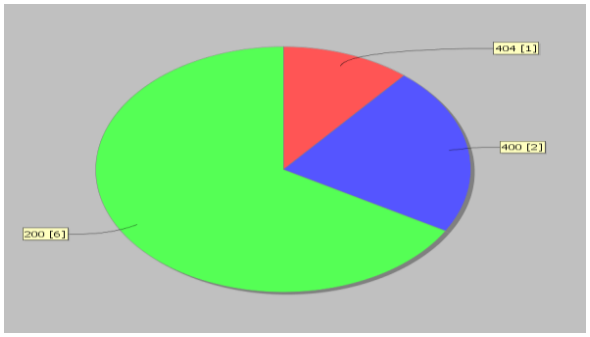


**Figure 6.Grouped Server Error Response Code with the number of occurrences.**

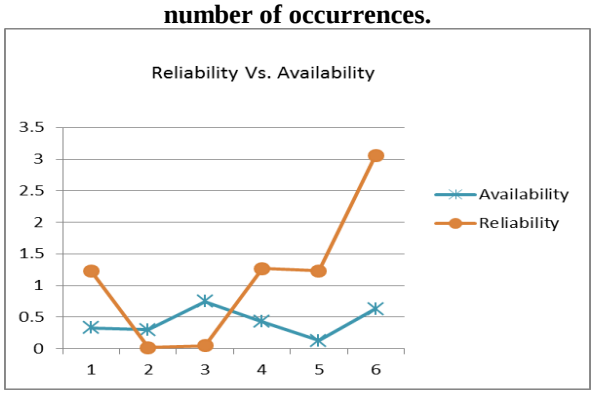


**Figure 7. Measure of Server Availability and Reliability for different Service Request.**

## V. CONCLUSION

By analyzing the unique problems and information sources for the web environment, an approach for identifying and characterizing web errors and for assessing and improving web site reliability based on information extracted from existing web logs is developed. Our results demonstrated that the error distribution across different error types and sources is highly uneven. In addition, missing file distributions, workload distribution, as well as reliability distribution for individual types of requested files are all quite uneven. Our analysis results can help web site owners to prioritize their web site maintenance and quality assurance effort and to guide further analyses, such as root cause analysis, to identify problem causes and perform preventive and corrective actions. All these focused actions and efforts would lead to better web service and user satisfaction due to the improved web site reliability.

## REFERENCES

[1] H. Elfawal Mansour, Member, IEEE, and T. Dillon,Member, IEEE, "Dependability and Rollback Recovery For Composite Web Services".IEEE Transaction On Services Computing, VOL.4, No.4.3, October-December 2011.

[2] Hirokazu Ozaki, Member, IEEE, and Atsushi Kara, Member, IEEE, "Reliability Analysis of M -for-N Shared Protection System With General Repair-Time Distributions".IEEE Transaction On Reliability, VOL. 60, NO. 3, September 2011.

[3] Miao Jiang, Member, IEEE, Mohammad A. Munawar, Member, IEEE, Thomas Reidemeister, Member, IEEE, And Paul A.S. Ward, Member, IEEE." System Monitoring with Metric Correlation Models." IEEE Transaction on Network and Service Management", VOL. 8, NO. 4, December 2011.

[4] Fatemeh Borran, Martin Hutle, Nuno Santos, and Andre Schiper, Member, IEEE, "Quantitative Analysis of Consensus Algorithms", IEEE Transaction On Dependable And Secure Computing VOL. 9, NO. 2, March/April[ 2012.

[5] Zheng Z, Lyu MR (2010) Collaborative reliability prediction of service-oriented systems. ICSE c610, May 2–8 2010, Cape Town, South Africa Copyright 2010 ACM 978-1-60558-719-6/10/05.

**P.Maruthurkarasi (alias) Rohini.** She received B.Tech. Information Technology degree of Anna University, Tiruchirapalli from Syed Ammal Engineering College, Ramanathapuram in April 2010. She is pursuing II year M.Tech IT Degree of Anna University from K.L.N. College of Information Technology during 2013. She is an active researcher in the field of web services. She published many papers in national and international conferences including papers listed in IEEE Xplore. This work is considered as partial fulfillment for the award of M.Tech (IT) Degree of Anna University during 2013.

**C.Jayapraksh,** He received MCA Degree of Madras University from St Joseph College of Engineering in April 1999. He received M.E. Computer Science and Engineering Degree of Satyabama University in May 2009. His research area includes Web Services and Distributed Computing. He is a research scholar of Satyabama University under Faculty of Computer Science & Engineering since June 2009. He has published many research papers in International Conferences and National Journals including some listed in IEEE Xplore.

**R.Balaji Ganesh,** He received B.E. Computer Science and Engineering degree of Anna University from P.S.R. Engineering College, Sivakasi in April 2006. He received M.Tech. Computer & Information Technology degree of Manonmaniam Sundaranar University in April 2011. His research area includes Image & Video Processing, Web Services, Data Mining. He has published many papers in International Conferences including some in European IADIS Conferences. He is a visiting faculty & research scholar of Manonmaniam Sundaranar University with active participation in CDAC INTRANSE project.