International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 2, February - 2014

Reliability Estimation and Analysis of DDL MYSQL Server by using Generalized Gamma and Weibull Distribution

M. Y. Haggag
Department of Mathematics,
Faculty of Science, Al-Azhar University,

Heba Nagaty Mohamed
Department of Mathematics
Faculty of Science, El fayoum University

***Abstract* - The time between failures for different Operating Systems (Windows and Linux) of DDL MYSQL open source data base server are analyzed and compared. The purpose of this study is to estimate and compare the reliability of two Operating Systems (Windows and Linux) of DDL MYSQL server by using Generalized Gamma and Weibull Distribution which are the best distributions in their rankings. In the result the Reliability Estimation of two Operating Systems are evaluated and compared theoretically and graphically.**

1. INTRODUCTION

Software reliability [1] is one of the important parameters of software quality. It is defined as the probability of failure-free software operation in a specified environment for a specified period of time. An earlier researcher as [3] has studied Reliability Estimation and Analysis of Linux Kernel and [4] has studied Estimation and Analysis of MYSQL Database Server Reliability using Beta and Generalized Gamma Distribution.

The Mysql database [5] has become the most popular open source database in the world because of its high performance, high reliability and ease of use. The purpose of this study is to compare between the DDL server database in operating systems (Windows and Linux), Where DDL (Data Definition Language) [2] is a language used by a database management system (like Mysql)that allows users to define the database and specify data types, structures and constraints on the data.

To find reliability two types of data can be used: time between failures and fault count. In case of "time between failures" the input parameter of study is the intervals of successful operations. A probability distribution model whose parameters are estimated by using appropriate mathematical technique reflects the pattern of these intervals. In case of "fault count "the input parameter of study is the number of faults in a specified period of time rather than the times between failures.

In this paper we discussed in the following sections: Section 2 provides some mathematical background of Reliability estimation. Section 3 concentrates on Bug Collection, Bug pre-processing and Bug analysis. Section 4 discussed Goodness of fit test and Parameter Estimation by using Maximum Likelihood Estimation .Section 5 show the proposed Methodology Used Weibull++ tool [7] for mathematical and statistical calculation. Section 6 shows the Reliability Evaluation of two Operating Systems. Finally conclusion and References of this paper are shown.

2. BACKGROUND

Software Reliability is defined as: the probability of failure-free software operation for a specified period of time in a specified environment. There are two approaches for prediction of Software Reliability: early stage where reliability estimated during design phase and later stage where reliability estimated during operational stage.
Software Reliability depends upon failure data of the software. Failure behaviour can be represented by various manners such as Probability Density Function (PDF) and Cumulative Distribution Function (CDF) which is derived from PDF and is given by equ. (1):

$$F(t) = P(x \le t) = \int_{-\infty}^{t} f(x)dx \qquad (1)$$

There for f(t) is the rate of change of F(t). If the random variable T denotes the failure time then F (t) is the probability that the system will fail by time t. Then F(t) is the unreliability function and R(t) is the Reliability function and given by equation (2):

$$R(t) = 1 - \int_{-\infty}^{t} f(x)dx \qquad (2)$$

Another function that can be derived from PDF is the failure rate function (Hazard Rate Function) which is defined by equation (3):

$$\lambda = \frac{f(t)}{R(t)} = \frac{f(t)}{1-\int_{-\infty}^{t} f(x)dx} \qquad (3)$$

3. BUG COLLECTION, BUG PRE-PROCESSING AND BUG ANALYSIS

The approach to the Reliability estimation of the two operating systems Linux and Windows consists of three steps:
1. Bug Collection: is associated with collecting failure data extracted directly from the following web site http://www.mysql.Bugs.org and Bugs of operating

system Linux are collected from 5/7/2004 to 19/8/2013 and Bugs of operating system Windows are collected from 14/4/2004 to 11/8/2013.

2. Bug preprocessing: in this step such noises are removed.
3. Bug analysis: the preprocessed data is stored in MYSQL database, where MYSQL is an open source database system.

Before applying Goodness of fit test on data collected for each operating system bug frequency corresponding to time to failure in month is plotted and show in Figure 1, Figure 2. The total of Bugs recorded for Windows are 39, and for Linux are 75.

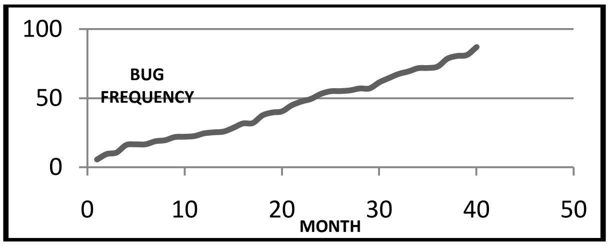

Figure 1. Monthly Bug Frequency of Windows

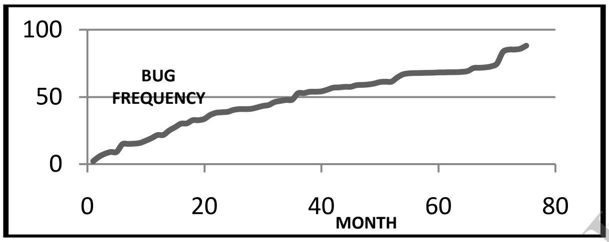

Figure 2. Monthly Bug Frequency of Linux

4. GOODNESS OF FIT TEST AND PARAMETER ESTIMATION

The Goodness of fit test is used to identify whether of the following distributions which are the most commonly used life distributions is suitable for collecting data or not.

- 1 and 2 parameter exponential distributions.
- 1, 2 and 3 parameter Weibull distributions.
- Normal distribution.
- Lognormal distribution.
- Generalized Gamma (G-Gamma) distribution.
- Logistic distribution.
- Log Logistic distribution.
- Gumbel distribution.

There are a lot of methods for The Goodness of fit test but the method of Maximum Likelihood Estimation is considered as the best method of Parameter estimation.

Maximum Likelihood Estimation

Maximum Likelihood Estimation [8] is used to estimate distribution parameters by maximizing the value of Likelihood function. This Likelihood function is based on the probability density function (PDF) for a given distribution, i.e. if (PDF) is f $(x_i, \theta_1, \theta_2, \ldots, \theta_k)$,where x represents the data (times-to-failure) and $\theta_1, \theta_2, \ldots, \theta_k$ are the parameters which is to be estimated. Then Likelihood function is given by equation (4):

$$L = \prod_{i=1}^{n} f\,(x_i, \theta_1, \theta_2, \ldots, \theta_k) \qquad (4)$$

Where n: is the number of failure data points. Then taking log-likelihood function which is defined by equation (5):

$$\ln L = \sum_{i=1}^{n} \ln f\,(x_i, \theta_1, \theta_2, \ldots, \theta_k) \qquad (5)$$

Finally, parameters are estimated by using the following partial derivatives given by equation (6):

$$\frac{\partial \ln L}{\partial \theta_j} = 0, j = 1,2, \ldots, k \qquad (6)$$

5. THE PROPOSED METHODOLOGY

In this research Weibull++ tool is used for mathematical and statistical calculation.

First: Parameters for all life data distribution are estimated by maximum likelihood estimation and presented in the following table (1) and table (2) for two operating system Linux and Windows respectively.

Second: Calculate Log-Likelihood Function By using parameters estimation and presented in the following table (3) and table (4) for two operating system Linux and Windows respectively.

Third: A distribution having maximum LKV is considered as best distribution fitted the given data.

Table 1 . Parameter Estimation for Linux

Distribution	Parameters
Exponential 1	μ =48.742639
Exponential 2	μ =46.8096,γ=1.933
Weibull 2	β=2.32103,η=54.57108
Weibull 3	β =6.167074,η=121.856165 ,γ=-64.2574
Normal	μ=48.74263,std=21.99186
Lognormal	mean=3.71399,Std=0.71288
G-Gamma	μ=4.222010,σ=0.255851, λ=2.583114
Gamma	μ=2.77007,k=3.05409
Logistic	μ=49.90922,σ=12.8883
Log-Logistic	μ=3.8324,σ=0.351556
Gumbel	μ=59.25009,σ=18.792409

Table 2. Parameter Estimation for Windows

Distribution	Parameters
Exponential 1	μ =43.56212
Exponential 2	μ =38.16212,γ=5.400
Weibull 2	β=1.96948,η=49.16702
Weibull 3	β =1.89676,η=47.79413 ,γ=1.305
Normal	μ=43.56212,std=23.41969
Lognormal	mean=3.58866,Std=.67598
G-Gamma	μ=4.28027,σ=0.226218, λ=3.640866
Gamma	μ=2.72656,k=2.85087
Logistic	μ=43.05427,σ=14.122369
Log-Logistic	μ=3.647028,σ=0.386143
Gumbel	μ=55.23509,σ=21.667332

Table 3 . Log-Likelihood Value for Linux

Distribution	LKV	Rank
G-Gamma	-332.51	1
Weibull 3	-335.93	2
Gumbel	-336.9	3
Normal	-337.7	4
Logistic	-339.9	5
Weibull 2	-340	6
Gamma	-347.1	7
Log-Logistic	-353	8
Lognormal	-359.08	9
Exponential 2	-363.4	10
Exponential 1	-366.4	11

Table 4 . Log-Likelihood Value for Windows

Distribution	LKV	Rank
G-Gamma	-178.6	1
Weibull 3	-180.5	2
Weibull 2	-180.79	3
Gamma	-181.67	4
Normal	-182.4	5
Lognormal	-184.14	6
Logistic	-184.28	7
Gumbel	-184.58	8
Log-Logistic	-184.68	9
Exponential 2	-185.6	10
Exponential 1	-190.9	11

From the previous tables (3), (4), it's clear that Generalized Gamma and Weibull Distribution is best suited and may be considered for reliability estimation.

We used the web site in [8] to determine the Generalized Gamma three parameters Distribution. We used the web site in [9] to determine the Weibull three parameters Distribution

A. Construction of Reliability Model using the Weibull Distribution

The Probability Density Function of Weibull Distribution is given by:

$$f(T) = \frac{\beta}{\eta}\left(\frac{T-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{T-\gamma}{\eta}\right)^{\beta}}, \beta > 0, \eta > 0 \qquad (7)$$

Where

β is the shape parameter, also known as the Weibull slope.

η is the scale parameter

γ is the location parameter

The cumulative Distribution Function of Weibull Distribution is given by:

$$F(T) = e^{-\left(\frac{T-\gamma}{\eta}\right)^{\beta}} \qquad (8)$$

The Reliability Function of Weibull Distribution is given by:

$$R(T) = 1 - e^{-\left(\frac{T-\gamma}{\eta}\right)^{\beta}} \qquad (9)$$

By substituting from Table (1), Table (2) of Parameter Estimation in the equations (7), (9), we get:

$$\begin{cases} f(T) = \frac{1.89676}{47.79413}\left(\frac{T-1.305}{47.79413}\right)^{.89676} e^{-\left(\frac{T-1.305}{47.79413}\right)^{1.89676}} & \text{(Windows)} \\ f(T) = \frac{6.167074}{121.85616}\left(\frac{T-(-64.2574)}{121.85616}\right)^{5.167074} e^{-\left(\frac{T-(-64.2574)}{121.85616}\right)^{6.167074}} & \text{(Linux)} \end{cases} \qquad (10)$$

and

$$\begin{cases} R(T) = 1 - e^{-\left(\frac{T-1.305}{47.79413}\right)^{1.89676}} & \text{(Windows)} \\ R(T) = 1 - e^{-\left(\frac{T-(-64.2574)}{121.85616}\right)^{6.167074}} & (Linux) \end{cases} \qquad (11)$$

B. Construction of Reliability Model using Generalized Gamma Distribution:

The Probability Density Function of Generalized Gamma Distribution is given by:

$$f(T) = \frac{\beta}{\Gamma(k).\theta}\left(\frac{T}{\theta}\right)^{k\beta-1} e^{-\left(\frac{T}{\theta}\right)^{\beta}}, \theta > 0, k > 0, \beta > 0 \quad (12)$$

Where

k *and* β are the shape parameter.

θ is the scale parameter.

Gamma Function has the formula:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

But, Weibull++ uses a reparameterization with parameters k, β and θ as shown in the following:

- $\mu = \ln(\theta) + \frac{1}{\beta}.\ln\left(\frac{1}{\lambda^2}\right), \mu > 0$ is the location parameter.
- $\sigma = \frac{1}{\beta\sqrt{k}}, \sigma > 0$ is the scale parameter.
- $\lambda = \frac{1}{\sqrt{k}}$ is the shape parameter. (13)

The Cumulative Function of Generalized Gamma Distribution is given by:

$$F(T) = \frac{\gamma\left(k, \left(\frac{t}{\theta}\right)^{\beta}\right)}{\Gamma(k)} \quad (14)$$

The Reliability Function of Generalized Gamma Distribution is given by:

$$R(T) = \frac{\gamma\left(k, \left(\frac{t}{\theta}\right)^{\beta}\right.}{\Gamma(k)} \quad (15)$$

By substituting from Table (1), Table (2) of Parameter Estimation in the equations (13), we get the following values in two operating systems (Windows and Linux) , respectively

$$K = .0754383, \beta = 16.094475, \theta = 84.8469 \quad \text{(Windows)}$$

$$k = 0.149869, \beta = 10.096149, \theta = 82.269615 \quad \text{(Linux)}$$

And then we get the following equations (16):

$$f(T) = \frac{\gamma\left(.0754383, \left(\frac{t}{84.8469}\right)^{16.094475}\right)}{\Gamma(0.0754383)} \quad \text{(Windows)}$$

$$f(T) = \frac{\gamma\left(.149869, \left(\frac{t}{82.269615}\right)^{10.096149}\right)}{\Gamma(0.149869)} \quad \text{(Linux)} \quad (16)$$

And

$$R(T) = 1 - \frac{16.094475}{\Gamma(.0754383)}\left(\frac{T}{84.8469}\right)^{-14.880335} e^{-\left(\frac{T}{84.8469}\right)^{16.094475}} \quad \text{(Windows)}$$

$$R(T) = 1 - \frac{10.096149}{\Gamma(0.149869)}\left(\frac{T}{82.269615}\right)^{-8.58304} e^{-\left(\frac{T}{82.269615}\right)^{10.096149}} \quad \text{(Linux)} \quad (17)$$

6. RELIABILITY EVALUATION:

It is clear from the goodness of fit section that best distribution appropriate for collected sample are Generalized Gamma Distribution with three parameters and Weibull Distribution with three parameters.

In this section the PDF and Reliability of Generalized Gamma Distribution with three parameters and Weibull Distribution are evaluated in the following tables (5),(6) by using equations (10),(11),(16),(17) ,and the corresponding graphs of the PDF and Reliability for each distributions are show in the following graphs (3),(4),(5),(6).

Table (5) The PDF of Weibull Distribution and G-Gamma Distribution for Windows &Linux

Month No.	Windows		Linux	
	f(t)-G-Gamma	f(t)- weibull	f(t)-G-Gamma	f(t)- weibull
5	0.008114516	0.003965289	0.0046843	0.002649
10	0.009412931	0.008275655	0.006685	0.003735
15	0.010266748	0.011784083	0.008231	0.00511
20	0.010919108	0.014449878	0.0095402	0.006786
25	0.01145353	0.016241065	0.0106974	0.008749
30	0.011909544	0.0171763	0.0117461	0.010935
35	0.012309227	0.017326563	0.0127111	0.013224
40	0.012666219	0.016805063	0.0136055	0.015426
45	0.012989341	0.015752378	0.0144304	0.017289
50	0.013283552	0.01432049	0.0151668	0.018526
55	0.013547534	0.012658195	0.015759	0.018864
60	0.01376302	0.010899465	0.0160854	0.018121
65	0.013862474	0.009155594	0.0159207	0.016285
70	0.013647303	0.007511239	0.0149169	0.013558
75	0.012631968	0.006023932	0.012689	0.010341
80	0.009967771	0.004726287	0.009142	0.007134
85	0.005317035	0.00362999	0.0049902	0.004388
88	0.002481168	0.003067554	0.0028354	0.003082
Average	0.011030166	0.010342088	0.011102	0.010346

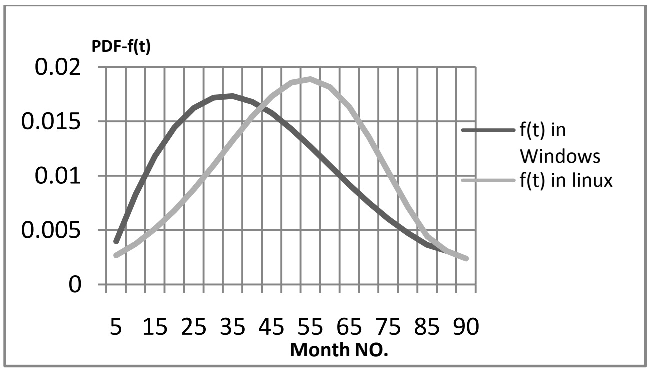

Figure 3. PDF of Weibull Distribution for Windows &Linux

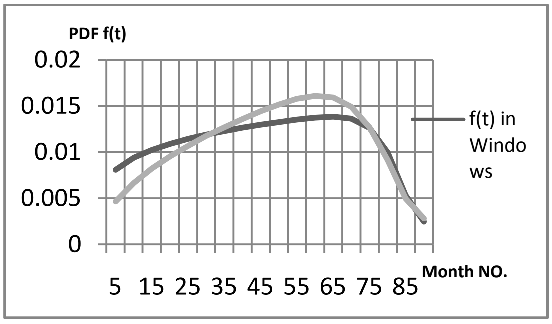

Figure 4. PDF-of G-Gamma Distribution for Windows &Linux

Table 6. The Reliability of Weibull and G-Gamma Distribution for Windows &Linux

Month No.	Windows		Linux	
	R(t)-G-Gamma	R(t)-weibull	R(t)-G-Gamma	R(t)- weibull
5	0.966583	0.99224524	0.984521	0.969795
10	0.922473	0.96130479	0.95582	0.953951
15	0.87316	0.91081577	0.918403	0.931965
20	0.820134	0.84486975	0.873899	0.90235
25	0.764164	0.76777766	0.823253	0.863623
30	0.705729	0.68388866	0.767105	0.814486
35	0.645162	0.59732532	0.705931	0.754098
40	0.582708	0.51174416	0.640111	0.68239
45	0.518556	0.43016062	0.56999	0.600403
50	0.452862	0.35485219	0.495953	0.510544
55	0.38577	0.28733889	0.418559	0.416648
60	0.317465	0.22842969	0.3388	0.323717
65	0.248327	0.17831797	0.258517	0.237271
70	0.179349	0.13670636	0.180988	0.162362
75	0.113162	0.10294244	0.111402	0.102512
80	0.0557885	0.07614977	0.0563443	0.058935
85	0.0169778	0.0553431	0.0210462	0.030398
88	0.00544105	0.04531424	0.00943495	0.019264
Average	0.476322853	0.43184461	0.507226525	0.492029

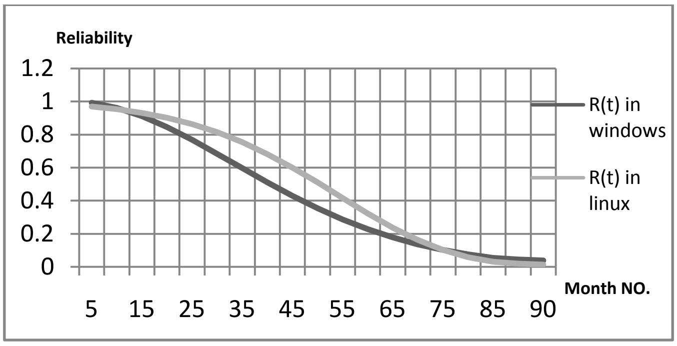

Figure 5. Show compare between the Reliabilities of Windows and Linux by using "Weibull Distribution"

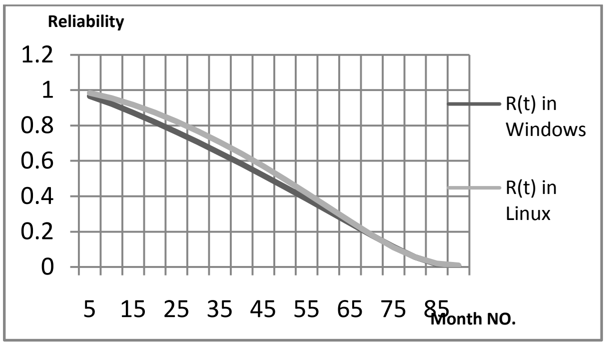

Figure (6) show compare between the Reliabilities of Windows and Linux by using "G-Gamma Distribution"

7. CONCLUSIONS:

In the above study a detail methodology to estimate reliability of two Operating System (Windows and Linux) are discussed and it has been analyzed by using two fitted Distributions: Weibull Distribution and G-Gamma Distribution.

The average value of Probability Density Function and Reliability of (Weibull Distribution and G-Gamma Distribution) for each Operating System are calculated, and it has shown that:

Operating system "Linux" is most Reliable than Operating system "Windows" for each two Distributions Weibull Distribution and G-Gamma Distribution.

REFERENCES:

1. Lyu, Michael R. "Handbook of software reliability engineering." (1996).
2. IEEE Reliability Society, "IEEE recommended practice on software reliability", IEEE Std 1633-2008, June 2008.
3. Sanjeev Kumar Jha, Dr. A.K.D.Dwivedi, Dr. Amod Tiwari." Reliability Estimation and Analysis of Linux Kernel"(2011) IJCST Vol. 2, Issue 2
4. Jha, Sanjeev Kumar, Pankaj Kumar, and A. K. D. Dwivedi. "Estimation and Analysis of MYSQL Database Server Reliability using Beta and Generalized Gamma Distribution". (IJCET) Journal Impact Factor 3.2 (2012):pp. 354-371.
5. Operating Systems Source of Bugs: http://bugs.mysql.com
6. F. S. G. RICHARDS, "A Method of Maximum-likelihood Estimation", http://www.jstor.org/pss/2984037
7. Weibull++, Reliability Function, [Online] Available: http://www.weibull.com and http://www.reliasoft.com
8. en.wikipedia.org/wiki/Generalized_gamma_distribution
9. http://reliawiki.org/index.php/The_Weibull_Distribution
10. H. Pham, Software Reliability. Springer Verlag, 2000.
11. Gavin E. Crooks (2010), The Amoroso Distribution, Technical Note, Lawrence Berkeley National Laboratory
12. Guo, Huairui, Jin, Tongdan, and Mettas, Adamantios. "Design Reliability Demonstration Tests for One-Shot Systems Under Zero Component Failure," IEEE Transactions on Reliability, Vol. 60, No. 1, pp. 286-294, (March 2011).