# STATISTICAL INTERPRETATIONS OF THE TURNAROUND TIME VALUES FOR A SCALABLE 3-TIER GRID-BASED COMPUTING ARCHITECTURE

Article · January 2013

6 authors, including:

Nureni Ayofe Azeez
University of Lagos
50 PUBLICATIONS   143 CITATIONS

SEE PROFILE

Philip Abidoye
University of South Africa
11 PUBLICATIONS   76 CITATIONS

SEE PROFILE

Ademola Adesina
University of the Western Cape
18 PUBLICATIONS   103 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Managing information in education  View project

Project   Cybersecurity  View project

# STATISTICAL INTERPRETATIONS OF THE TURNAROUND TIME VALUES FOR A SCALABLE 3-TIER GRID-BASED COMPUTING ARCHITECTURE

N.A. Azeez*[1]; A.P. Abidoye[1]; A.O. Adesina[1]; K.K. Agbele[1]; I.M. Venter[1]; and A.S. Oyewole[2]

[1]Department of Computer Science, University of the Western Cape
Bellville, Cape Town, South Africa

[2] Department of Computer Science, Durban University of Technology, Durban, South Africa.
E-mail: nurayhn@yahoo.ca*

*Abstract*

*The concept of scalability for the survival and full-scale implementation as well as efficient utilisation of any distributed system, particularly grid computing, cannot be over-emphasised. It is through this security concept that inter-domain and intra-domain resource sharing, distribution and aggregation can be adequately ensured and guaranteed. This paper is an extension of a paper titled "Towards achieving scalability and interoperability in a triple-domain grid-based environment (3DGBE)" presented at the Information Security for South Africa (ISSA), conference, 2012. The objective of this paper is to provide a comprehensive statistical analysis of the values obtained for the turnaround times for the three domains observed against the number of grid nodes when scalability was being evaluated using ANOVA. Further results obtained also give a summary of the statistical analysis of the values obtained with respects to the three domains vis-à-vis the mean ($\bar{X}$), standard deviation and the variance using ANOVA. The essence of the statistical interpretation of these values is to affirm and confirm the significance difference in the mean of group of the three domains considered with respect to variation in the number of grid nodes as well as the number of service requesters as presented.*

*Keywords: statistical, 3-tier, grid computing, scalability, interpretation, ANOVA*

## Introduction

In the previous article titled "Towards achieving scalability and interoperability in a triple-domain grid-based environment (3DGBE)" three basic challenges affecting the full scale optimization of grid computing were addressed. The challenges are scalability, interoperability and efficient access control.

To tackle these problems, a novel architectural model built on three domains was designed. The architectural model has an autonomous local security monitoring unit as well as central security monitoring unit for the entire grid system.

To tackle these problems, a novel architectural model built on three domains was designed. The architectural model has an autonomous local security monitoring unit as well as central security monitoring unit for the entire grid system.

- Convincing results in terms of scalability were achieved when the throughput and number of grid nodes as well as when the average turnaround was measured against the number of grid requesters. In order word, the effects of throughput both on the grid nodes and grid service requesters show that the 3DGBE is scalable.
- The results obtained in terms of interoperability when the operating systems, grid middleware, LSMU and CSMU as well as database were implemented and experimented with, proved that the model's framework is interoperable. There is a smooth

correspondence for message-passing between the LSMU and CSMU. Grid middleware integration with different operating systems across the domains also provide an interoperable platform for resource distribution.

- Finally, the efficient access control was assessed with hierarchical role based access control and experimented with a health scenario. The result obtained has clearly the benefit of efficient access control for the architecture designed.
- Based on the results obtained, the architectural framework has proved to be scalable when the average turnaround time was measured against the number of grid nodes.

In this paper, statistical evaluation of values obtained in terms of scalability is presented. The objective of this is to have a clear-cut of the summarised and succinct interpretations of some these values by drawing statistical inferences.

*Scalability is the ability of a grid system to act efficiently, sufficiently and adequately in handling both a small and large number of nodes. This includes the processing capacity configuration in order to achieve a secured multi-organisation resource-sharing environment* (Azeez, 2013).

Scalability as explained in (Azeez & Venter, 2012) provides comprehensive information on the need for provision for future expansion in grid entities, ranging from growth and increase in resources, domains as well as people participating on resource sharing across the grid based environment. To avoid confusion, for the first reader on what was presented at the Information Security for South Africa (ISSA), conference 2012, the architectures are summarily presented in Figures 6 and 7 without further and explicit explanation. However, for detailed explanation and a follow up on the discussion about the architectures, it is strongly recommended that a paper titled "Towards achieving scalability and interoperability in a triple-domain grid-based environment (3DGBE)" could be downloaded from the IEEE *Xplore* digital library, since the objective of this paper is to present the statistical evaluation and interpretation of the values obtained for the turnaround time for Domains A, B and C when the number of grid nodes and grid service requesters were varied accordingly.

*Average turnaround time (sec) is the time taken between the period of request for accessing resources in a 3DGBE and the return of detail request output to the grid user* (Azeez, 2013).

The values obtained for the turnaround time are summarized in Tables 1 and 2 when measured against the number of grid nodes and service requesters respectively. The percentage of the turnaround values per each domain are also presented as well as the *total turnaround time* for the three domains.

According to Rossiter, D G (2006), statistical analysis means aggregation of various approaches to process and evaluate large volume of data along with related information as well as its final report interpretation in order to grab better understanding of the data. This approach provides unbiased report and analysis of how unfamiliar event is based on used data.

In this paper, attempt was made to compare and contrast the significance difference in the mean, mean of mean, variance and the standard deviation for the values obtained for the turnaround time when the number of grid nodes and number of grid service requesters were varied.

**The means, variance and the standard deviations of the experiments**

**A.   Average turnaround time versus number of grid nodes**

To find the mean and the standard deviation as shown in Tables 1 and 2, the following two definitions are provided:

1. Let $M_{SB}$ denotes variance between the three domains considered
2. Let $M_{SW}$ denotes variance within the three domains considered

To evaluate both the means and standard deviation of the experiment shown in Figure 4, we construct hypothesis test based on the values obtained using ANOVA.

$H_O$: $\mu = \mu A = \mu B = \mu C$, where A, B and C are domains considered. (There is no difference in the mean of the three Domains)

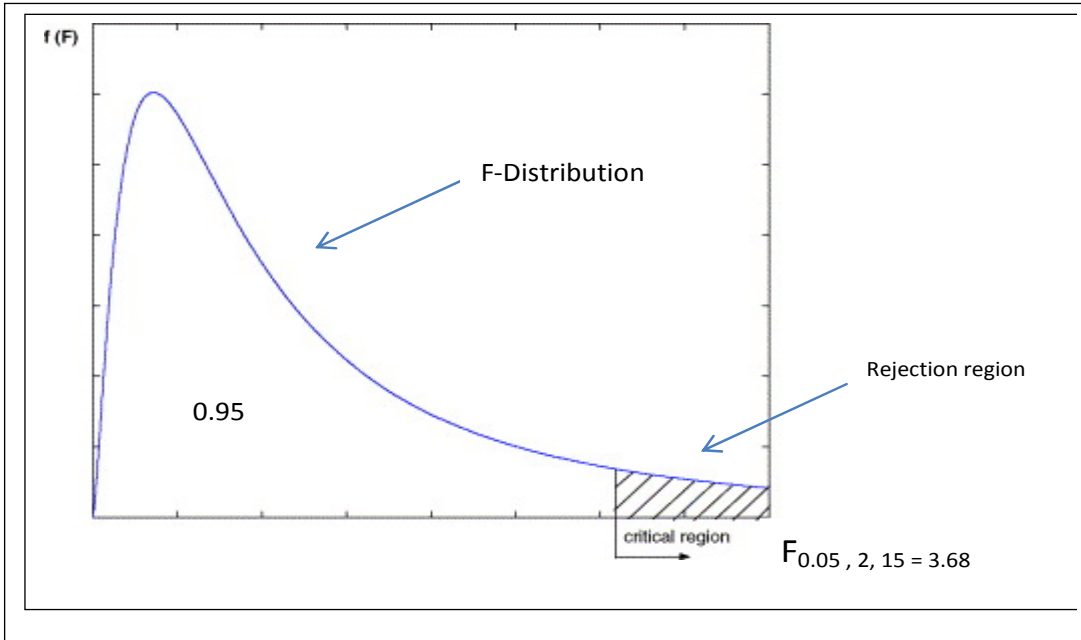$H_1$: At least one of the mean is different from the others.



*Figure 1: Showing values of 3.68 at* $\mathbf{F_{0.05\,,\,2,\,15}}$

It is noted that there are currently the value of K = 3 domains, that is, Domains A, B , C. Therefore, $DoF_N$ = K- 1 = 3-1 =2. The sum of data for all the three domains denoted as N = n1 + n2 + n3 = 6 + 6 + 6 = 18.

Using the $DoF_D$ = N – K = 18 -3 = 15 and α = 0.05 (the Least Significant value).

The critical value if $F_{0.05\,,\,2}$, 3.68 (determined using F-Distribution table)

There is need to find : $\bar{\bar{x}}$ = mean of mean = $\bar{\bar{x}}$ =( $\bar{X}_A + \bar{X}_B +\ldots\ldots \bar{X}_N$) /K

, $M_{SB} = \sum ni\,(\bar{X} - \bar{\bar{x}})^2$ /K-1 and $M_{SW} = \sum (ni - 1)S_i^2$ /K-1

The mean of mean denoted as $\bar{\bar{x}}$ was determined as follows:

$\bar{\bar{x}}$ = 31.50 + 39.10 +43.40 /18 = 6.33,

The mean for each of the domains are evaluated as follows: $\bar{X}_{Domain\,A} = \sum X$ /n, = 31.5/ 6 = 5.25, $\bar{X}_{Domain\,B} = \sum X$ /n = 39.1/6 = 6.51 and $\bar{X}_{Domain\,C} = \sum X$ /n = 43.40/6 = 7.233

$s^2_{Domain\,A = 1/N} \sum X_i^2$ - $\bar{X}^2$ = 236.65/6 - $5.25^2$ = 11.879, while $s^2_{Domain\,B}$ = 14.73 and $s2_{Domain\,C}$ is 10.59

Mean of Mean ; $\bar{\bar{x}}$ = (31.5 + 39.1 + 43.4) / 18 = 6.33

Table 1: The values for average Turnaround time and the No of grid nodes for Domains A, B, C

| Parameters Determined | No of grid nodes | Turnaround time (T) (Domains A,B, C) | | | Total Turnaround time for Domains A,B, C | Average Turnaroun d time for |
|---|---|---|---|---|---|---|

| | | Domain A | Domain B | Domain C | | Domains A,B, C |
|---|---|---|---|---|---|---|
| | 10 | 12  = 30.77% | 14 = 35.89% | 13=33.33% | 39 | 13 |
| | 20 | 7.0 =29.17% | 8.0 =33.33% | 9.0=37.50% | 24 | 8 |
| | 30 | 4.2=23.33% | 6.4=35.55% | 7.4=41.11% | 18 | 6 |
| | 40 | 3.8=25.33% | 4.8=32.00% | 6.4=42.67% | 15 | 5 |
| | 50 | 3.1=25.83% | 3.9=32.50% | 5.0=41.67% | 12 | 4 |
| | 60 | 1.4=23.33% | 2.0=33.33% | 2.6=43.33% | 6.0 | 2 |
| $\sum X$ | ⟶ | 31.5 | 39.1 | 43.4 | | |
| $\bar{X}$ | ⟶ | 5.25 | 6.51 | 7.233 | | |
| $s^2$ | ⟶ | 11.879 | 14.73 | 10.59 | | |
| n | ⟶ | 6 | 6 | 6 | | N = 18 |

Also, from Table 20 shown, $M_{SB} = \sum ni\ (\bar{X} - \bar{\bar{x}})^2$ /K-1 could be determined as follows:

$$=[6(5.25 - 6.33)^2 + 6(6.51 - 6.33)^2 + 6(7.233 - 6.33)^2] / 2$$

$$= 6.042$$

Also, $M_{SW} = \sum (ni - 1) S_i^2$ /N-K

$$= [(6-1)\ 11.879 + (6-1)\ 14.73 + (6-1)\ 10.59] / 15$$

= 12.399, therefore, the test statistics is F = $M_{SB}$ / $M_{SW}$ = 6.042 / 12.339 = 0.4872
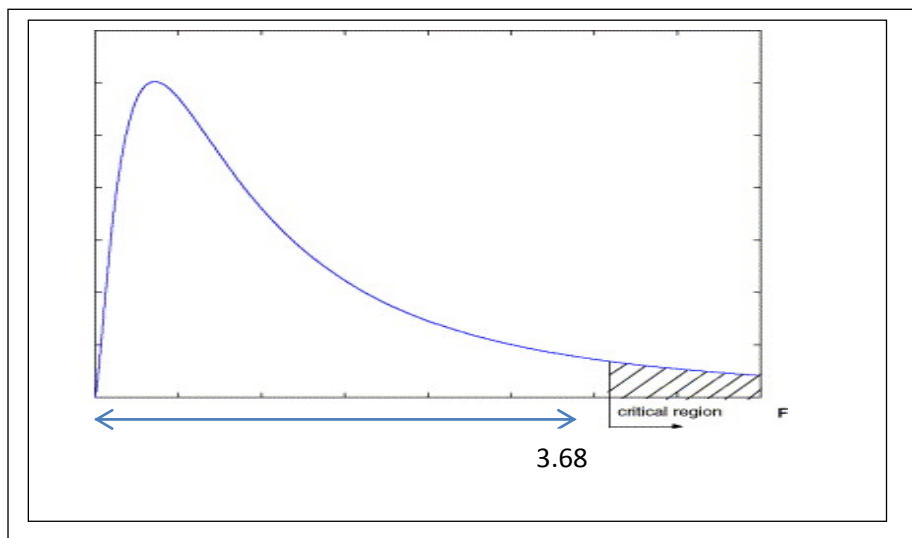


*Figure 2: Showing F-Distribution table for 3.68*

B. **Average turnaround time versus number of grid service requesters**

To find both the mean ($\bar{X}$) and the standard deviation ($s^2$ ) of experiment whose result is graphically presented in Figure 5, hypothesis was conducted based on the data obtained during experimentation.

As usual, the values of K, $DoF_N$ , N and $DoF_D$ are determined as : K = 3, $DoF_N$ = K – 1 = 3-1 = 2 , N = n1 + n2 + n3 = 5 + 5 + 5 = 15 while $DoF_D$ = N – K = 15 – 3 = 12 and α = 0.05 (the Least Significant value).

Table 2: The values for average Turnaround time and the No of grid service requesters for Domains A, B, C

| Parameters Determined | No of service requester | Turnaround time (T) (Domains A,B, C) | | | Total Turnaround time for Domains A,B, C | Average Turnaround time for Domains A,B, C |
|---|---|---|---|---|---|---|
| | | Domain A | Domain B | Domain C | | |
| | 50 | 1.1 = 36.66% | 1.4 = 46.66% | 0.5 = 16.66% | 3.0 | 1.0 |
| | 100 | 2.6 =30.95% | 3.1 = 36.90% | 2.7 = 32.14% | 8.4 | 2.8 |
| | 200 | 2.8 = 29.17% | 3.4 = 35.41% | 3.4 = 35.42% | 9.6 | 3.2 |
| | 300 | 3.0 = 29.41% | 3.6 = 35.29% | 3.6 = 35.29% | 10.2 | 3.4 |
| | 400 | 3.6 = 30% | 4.4 = 36.67% | 4.0 = 33.33% | 12.0 | 4.0 |
| | | | | | | |
| $\sum X$ | $\longrightarrow$ | 13.1 | 15.9 | 14.2 | | |
| $\bar{X}$ | $\longrightarrow$ | 2.62 | 3.18 | 2.84 | | |
| $s^2$ | $\longrightarrow$ | 0.6896 | 0.9776 | 1.5464 | | |
| n | $\longrightarrow$ | 5 | 5 | 5 | | N = 15 |

For the experiment whose values appear in Table 21, the critical value $F_{0.05}$ , 2, 3.89 (using F-Distribution table) is determined.

$s^2_{Domain\ A}$ = $1/N \sum X_i^2$ - $\bar{X}^2$ = 37.77/5 – (2.62)$^2$ = 0.68644, $s^2_{Domain\ B}$ = 0.9776 and $s^2_{Domain\ C}$ = 1.5464

To calculate the Mean of Mean ($\bar{\bar{x}}$) based on the values provided in the Table 21, therefore,

$\bar{\bar{X}}$ = 13.1 + 15.9 + 14.2 / 15 = 2.88

Also, from Table 2 shown, $M_{SB}$ = $\sum ni\ (\bar{X} - \bar{\bar{x}})^2$ /K-1 could be determined as follows:

= [5 (2.62 – 2.88 )$^2$ + 5(3.18 – 2.88)$^2$ + 5(2.84 – 2.88)$^2$]/2 = 0.796 / 2 = 0.398

Also,

$M_{SW}$ = $\sum ( ni - 1 ) S_i^2$ /K-1 was obtained as follows:

= [(5-1) 0.6896 + (5-1) 0.9776 + (5-1) 1.5464]/ 12

= 12.8544 / 12

= 1.0712

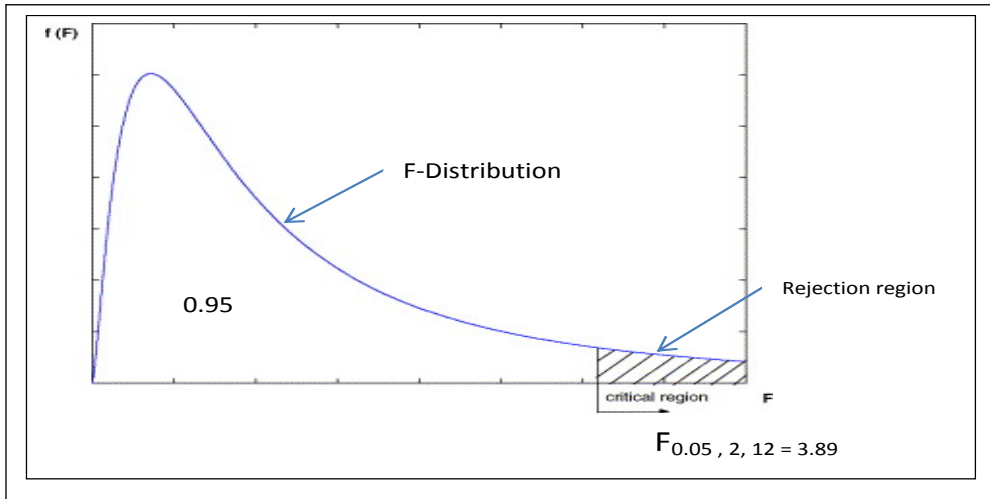Therefore, the statistics F =, $M_{SB}$ / $M_{SW}$

= 0.398 / 1.0712 = 0.3715

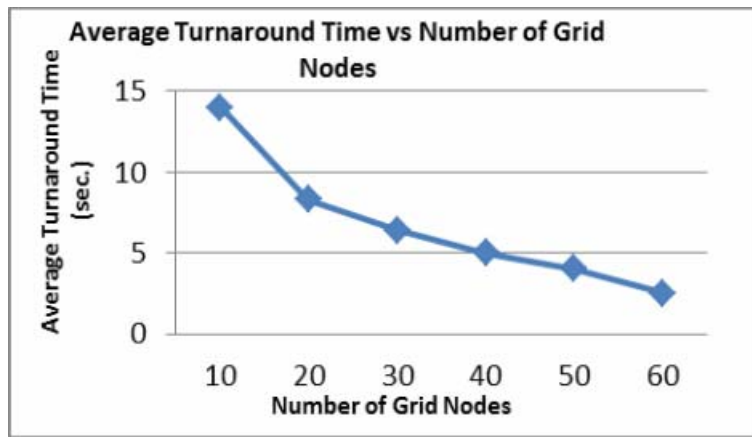*Figure 3: Showing values of 3.89 at* $\mathbf{F_{0.05\,,\,2,\,12}}$



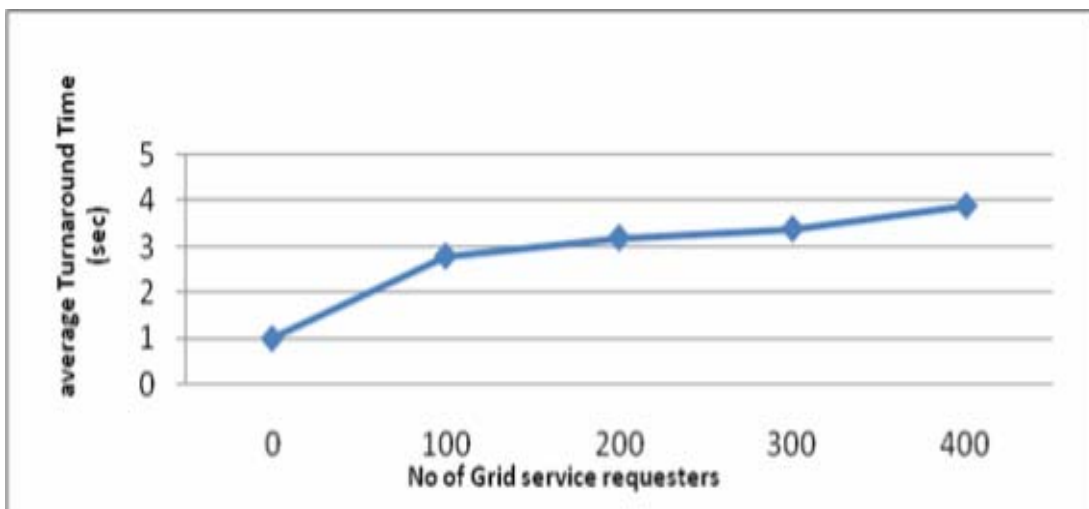Figure 4: Average turnaround time versus number of grid nodes



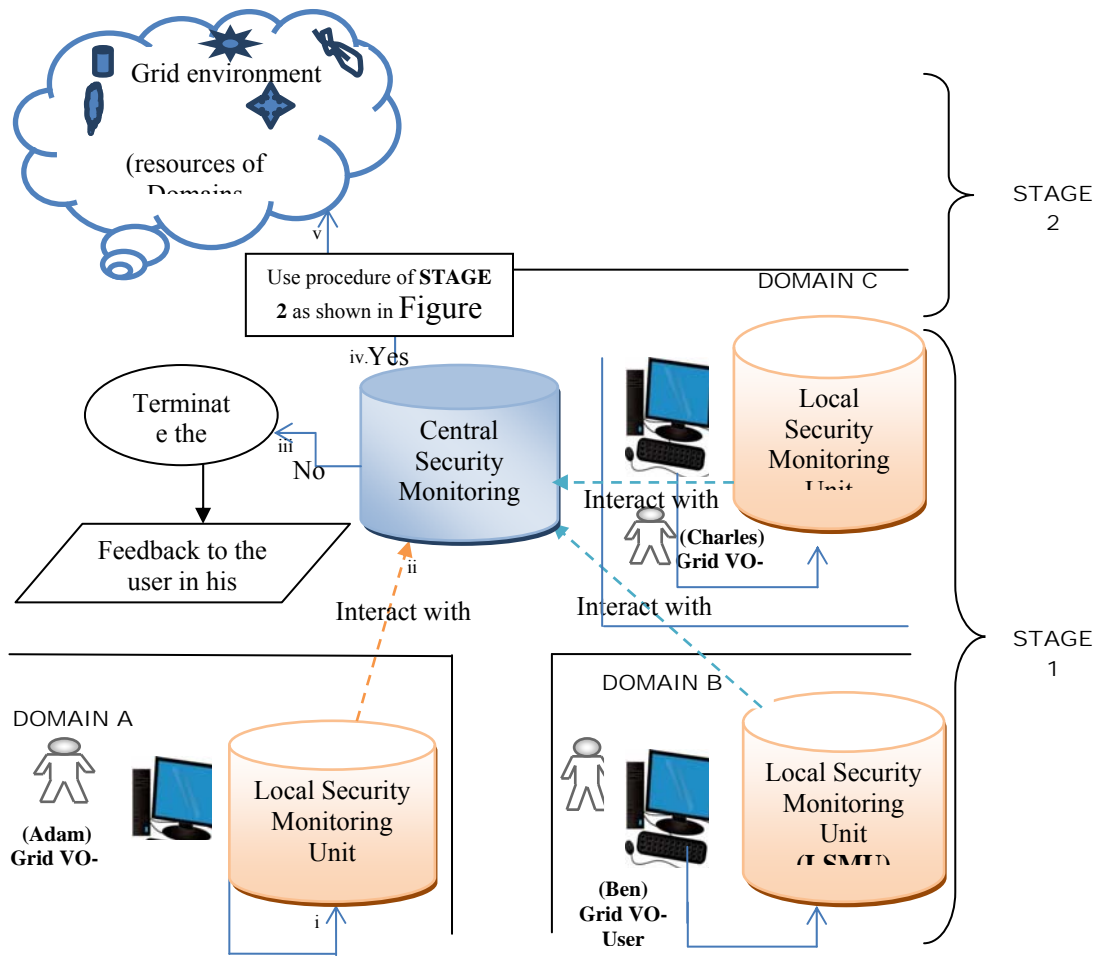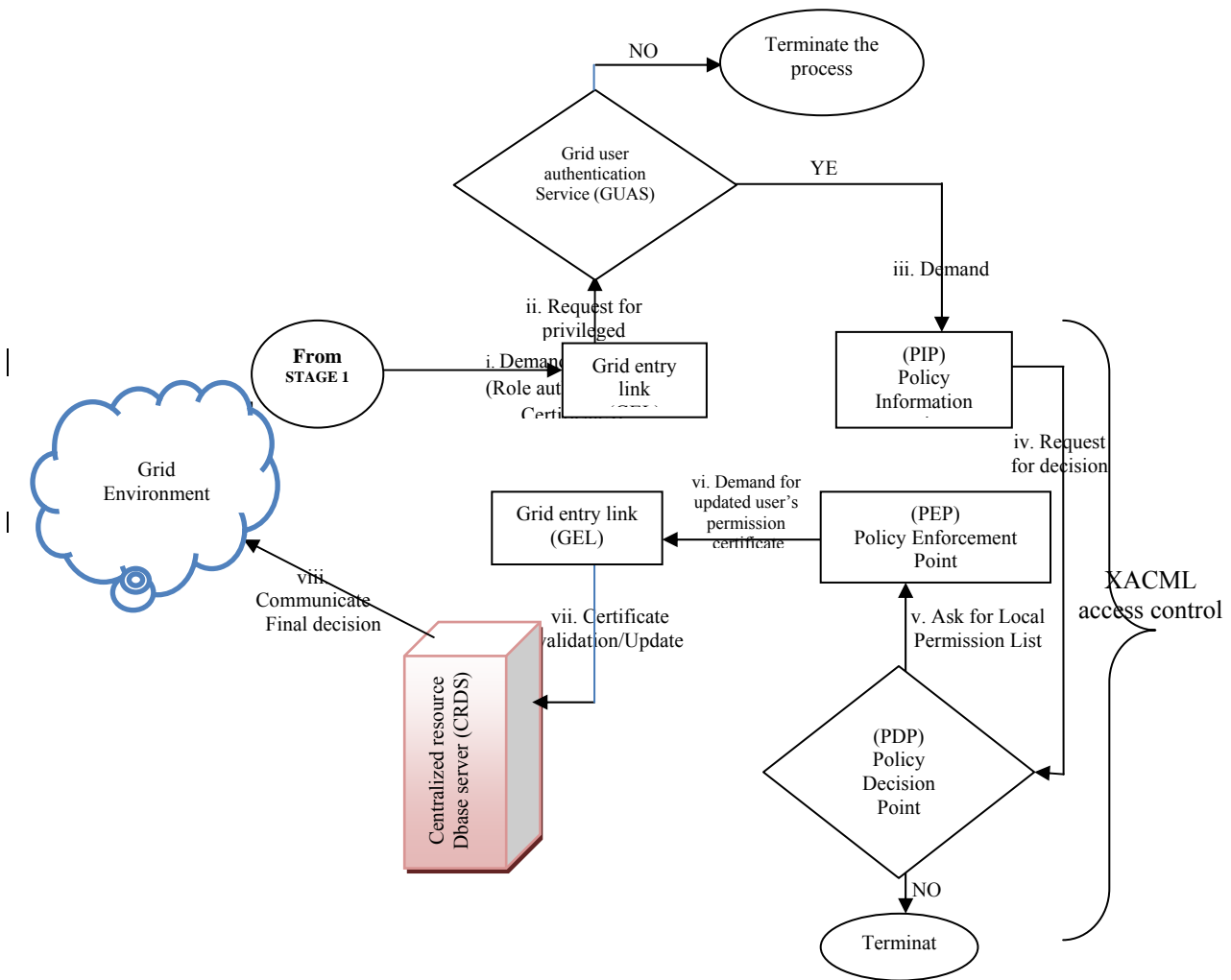*Figure 5:  Average turnaround time versus number of service requesters*

Figure 6: STAGE 1 of 3DGBE architectural framework of the proposed model
*(showing interaction between users, CSMU and LSMU)*

**Figure 7:     STAGE 2 of 3DGBE architectural framework of the proposed model**

**Conclusion**

**Deduction 1**

Since F-Statistical table falls to the left of F-distribution (0.4872 < 3.68) therefore, we accept the $H_O$. There is significantly no difference in the means of the three domains A, B and C even though the values of the grid nodes were varied independently across each of the domains as well as the corresponding values of the turnaround time. The graph of the Number of grid nodes with their corresponding average turnaround time is presented in Figure 4 based on the values analysed in Table 1.

**Deduction 2**

Finally, since F-Statistical table falls to the left of F-Distribution (0.3715 < 3.89), therefore , we accept  $H_O$.  We finally conclude that there significantly no difference in the *Mean* of group of Domains A, B, C. There is significantly no difference in the *Mean* even though, the values of number of grid service requesters were varied independently across each of the domains and the corresponding values of the average turnaround time. The graph is presented in Figure 5 using the values in Table 2; where the average turnaround time is plotted against the number of grid service requesters.

**References**

1. N.A. Azeez & I.M Venter (2012) "Towards achieving, scalability and interoperability in a Triple-Domain Grid-Based Environment" 11th Annual Information Security South Africa Conference ISSA 2012 , Johannesburg , South Africa, 15 – 17 August 2012 (Paper 26) (ISBN-978-1-4673-2158-7, IEEE Catalog Number: CFP1266I-CDR)
2. Azeez, N.A. (2013). PhD Thesis (Doctor of Philosophy). In Azeez, Towards ensuring scalability, interoperability and efficient access control in a triple-domain grid-based environment. University of the Western Cape, Bellville, South Africa.
3. Detsch, A., Gaspary, L., Barcellos, M., & Cavalheiro, G. (2004). Towards a Flexible Security Framework for Peer-to-Peer based Grid Computing. *2nd Workshop on Middleware for Grid Computing* (pp. 52-56). Toronto, Canada: ACM.
4. Rossiter, D.G (2006). An introduction to statistical analysis Overheads, Department of Earth Systems Analysis International Institute for Geo-information Science & Earth Observation (ITC). http://www.itc.nl/~rossiter/teach/stats/sintro.pdf

[i]Figures 4, 5, 6 and 7

---

[i] The details regarding Figures 4, 5, 6 and 7 can be found in NA Azeez & IM Venter (2012) as shown in the publication that appears in the first reference.