Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2018, 5(5):344-352**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Evaluating Quality of Service in Grid Computing Environment

Olufemi Ayodeji Odeniyi*, Taiwo F. Igbaroola, Afiss E. A. Kareem, Bukola Oyeladun Makinde, Nurudeen Lawal

Department of Computer Science, Osun State College of Technology, Esa-Oke, Nigeria

Abstract Grid Computing affords its users the opportunity to find and share data. However, due to the variation of data and the presence of large amount of data on the grid, some of the data do not get to their destination completely in real time thereby reducing the quality of service delivered unto the users. This work therefore addresses this challenge by dividing data into packets and then evaluates its performance in data transfer. Four quality of service metrics were adopted and simulated in the MATLAB simulation environment namely, one-way delay, bulk transport capacity, packet loss ratio and Internet protocol packet delay variation metrics. The outcome of the simulations was translated into graph to reveal the details which shows that the presence of these metrics in data transfer on the grid will increase or decrease the quality of service delivered to the grid users and that breaking up of data is efficient in data transfer.

Keywords Grid Computing, Resource Sharing, QOS Metrics, Data Transfer, Packets

1. Introduction

Grid Computing is a special type of distributed computing, that relies on complete computer (with on-board) CPU, storage, power supply, network interface, etc. connected to a network (private, public or the internet) by a conventional network interface, such as Ethernet. Grid Computing and the utilization of the global Grid infrastructure have presented significant challenges at all levels including conceptual and implementation models, application formulation and development, programming systems, infrastructures and services, resource management, networking and security have all led to the development of global research community.

In light of this, Grid Computing is concerned with applying the resources of many computers in a network to solve a single problem. Akinyemi *et al.*,[1] defined Grid computing as a new paradigm of distributed computing that facilitates virtual collaboration and interaction through the sharing of both software and hardware resources within a virtual organization. There are three reasons why Grid Computing is appearing to be a promising trend in the future [2]: Its ability to make cost-effective use of a given amount of computer resources; Parallel processing capacity and its ability to harness and manage the resources of many computers toward a common objective.

Therefore, Grid is the computing and data management infrastructure that provides the electronic underpinning for a global society in business, government, research, science and entertainment [3-6]. A grid is indeed an aggregate of software and hardware resources that are available for one another for the rendering of mutual non-trivial services [1]. Its infrastructure provides us with the ability to dynamically link together resources to ensemble and support the execution of large-scale, resource-intensive and distributed applications [6].

Grid Computing can be differentiated from all distributed computing paradigms [7] by a defining characteristic: efficient and optimal utilization of a wide range of heterogeneous, loosely coupled resources in an organization tied to sophisticated workload management capabilities or information virtualization, enabling resource sharing and dynamic allocation of computational environments composed of components from different domains, thus increasing access to data, promoting operational flexibility and collaboration, and allowing service providers to

efficiently scale to meet variable demands. The performance of these characteristics depends on the quality of network connections as network connects resources on the grid. However, parameters such as one-way delay, bulk transport capacity, packet loss ratio and Internet protocol packet delay variation, and so on are important to describe network performance to guarantee quality of service.

As Grid Computing increasingly enters the commercial domain, performance and Quality of Service (QOS) issues, such as customer's observed response times and throughput, are becoming a major concern. The inherent complexity, heterogeneity and dynamics of Grid computing environments pose some challenges in managing their capacity to ensure that QOS requirements are continuously met. Also, rapid advancement of communication technology has changed the landscape of computing. New model of computing, such as business-on-demand, web services, peer-to-peer networks, and grid computing have emerged to harness distributed computing and network resources to provide powerful services. The non-deterministic characteristics of the resource availability in these new computing platforms raise an outstanding challenge: how to support quality of service to meet a user's demand.

Quality of service (QOS) is the ability to provide different priority to different applications, users, data flows, or to guarantee a certain level of performance to a Data flow (for example, a required bit rate, delay, jitter, packet dropping probability and/or bit error rate). Metrics in the networking area, specifically to QOS, where a metric is defined as a quantitative value about any network aspect that permits studying its behaviour [9]. QOS metrics/parameters are used in ensuring an application or a service's performance. Different types of applications have applicable metrics – each important within its own space [9-11]. Braun and Staub [12] considered specifically QOS metrics such as Packet Delays (Queuing delay, Transmission delay, Propagation delay, One-Way Delay, Processing delay); Bandwidth metrics; Packet Loss Ratio. Also, Serral-Gracia *et al.,* [8] considered QOS metrics such as Connectivity metrics; IP Packet Delay, Delay Variation; Bulk Transport Capacity; Call Blocking metrics, Set-up Latencies, Bandwidth, Jitter, Packet Loss, among others, on the network level and user level as well for example mean opinion scores. Compute intensive applications would want to know about strength of processing power and memory usage; database applications would want to measure time per query or queries per second; environment with emphasis on the Grid network packets performance. The ability to deliver non-trivial quality of service is one of the three components to grid computing, along with coordinating resources that are not under centralized control and using standard, open, general purpose protocols and interfaces [13].

Furthermore, Quality of service measurement is a key to ensuring an application or a service's performance, however most approaches to QOS measurement have been focusing on that of resources optimization and network QOS aspect of the Grid are neglected, the networks are very important to connect data on the Grid. In data transmission on the grid, there are some difficulties experienced such as delay or loss of data due to failure on the part of the network connections. Monitoring the transmission of data along the network path from source to destination poses a challenge because of the lack of model to formalize this and due to the high volume of data to be transferred. This study presents a model to solve the difficulties experienced in data transmission by subdividing the data into packets and evaluating its behaviour from the source to the destination. In this model, the packets were evaluated using the QOS metrics. In other words, this study develops an appropriate QOS metrics for the Grid computing environment, concentrating on the evaluation of the behaviour and performance of network packets in Grid data transfer using the Quality of Service (QOS) parameters/metrics.

In addition, the Data Grids which serves as a database on the grid system needs to ensure reliable data transfer through data validation services set up to guarantee quality of service. For File-based Data Grids, there are ongoing efforts to provide a reliable file transfer service for the Grid [14]. This study therefore helps to intensify the effort by putting forward the concept of data being broken up in to smaller units called packets. As a result of large volume of data on the grid and due to the presence of geographically distributed users and machines, there arises the need to divide the data into packets. This is to ease the transfer of data along the network path and to ensure reliable data transfer. The packets which travel independently of one another are marked with the sender's address, destination address, and other pertinent information, including data about any errors introduced during the transfer, when the packets arrive at the receiving computer, they are reassembled. This study focuses

on resource layer of the grid-layered architecture because it is at the resource layer that access to data and access to computers are granted. However, Packet addressing and scheduling is out of the scope of this study.

2. Previous Research

The study of QOS metrics no doubt has been in existence for quite a long time. The following are different efforts in the application of QOS metrics for performance evaluation.

Simulation for evaluating the performance of MANETs (Mobile Ad-hoc Networks) was presented by Setty *et al*., [15]. The paper was subjected to the on demand routing protocol AODV and evaluated its performance in three different environments namely Random, Grid and Uniform. Four QOS metrics were investigated namely Average jitter, Average end-to-end delay, Packet delivery ratio and Throughput, in various simulation scenarios. The performance of AODV was evaluated by keeping the network speed and pause time constant, their work is different from this research, in the sense that simulation was performed for the Grid computing environment as well as the evaluation of packets used in data transmission in order to ensure that reliable data and quality of service is guaranteed to the Grid users.

In Quality of Service Support for Grid Computing, Colling *et al*., [16] proposed a framework for supporting Quality of Service guarantees via both reservation and discovery of best-effort services based on the matchmaking of application requirements, as well as Quality of Service performance profiles of the candidate services, different stages in the process of completing a typical Web services-based client-server interaction, their work is similar to this work but with emphasis on packets transmission from the client to the server using QOS metrics and evaluating its efficiency.

In Quality of Service Aspect and Metrics in Grid Computing, [17] investigated some issues that must be considered in designing Grid application that deliver appropriate QOS which include definition of metrics, relationship between the resource allocation an SLAs and QOS mechanism. An Insurance Company scenario(IC) was used to illustrate the types of QOS metrics (latency, throughput and availability) and the layer of the metrics in the Grid architecture. The user of the IC web portal will be able to request for three types of quotes. Similarly, to this work they argued that different types of QOS Metrics interact with one another on the Grid layers, other aspects of QOS including the use of optimization problem to find the allocation of cycles to compute resources that minimize the execution time were also considered, while satisfying the cost constraint. The approach presented in this work differs from theirs and focuses on evaluating the performance of packets in the grid network.

Kalogeraki *et al*., [18] reported that dynamic scheduling of distributed method invocations consists of algorithm that monitors the computation times and resource requirements of a job to determine a feasible schedule of method invocations on processors. Such a schedule is driven by the laxity and the priority of the job, although, scheduling is out of the scope of this work but the approach presented here serves as a model to predict the possible time of a data will take to reach its destination in Grid data transfer. Also, Kalogeraki *et al*., [18] stated that Data Management and Transfer in High-Performance Computational Grid Environments focuses on the fundamentals of Data Grid services, namely, secure, reliable, efficient data transfer and the ability to register, locate, and manage multiple copies of datasets; and presented the design and implementation for the Grid FTP Protocol to facilitate data transfer and the Globus replica management architecture.

Allcock *et al*., [14] reported that the GridFTP protocol implements extensions to FTP that provide GSI security and parallel, striped, partial, and third-party transfers, while the Globus replica management architecture supports the management of complete and partial copies of datasets. Preliminary performance measurements of the GridFTP prototype were also demonstrated. Their work is similar to this work because it involves data transmission but differs because GridFTP alone does not guarantee a level of quality of serviceand so there is the need to still subdivide the large data into packets and evaluates its behaviour and performance.

Hence, in the previous research reported above, some QOS metrics used are not addressing the splitting up of data and there arises a need for a model to evaluate the behaviour and performance of packets involved in data transmission in the Grid computing environment. Moreover, the quality of the packets employed will determine the quality of data that will be delivered to the Grid user.

3. Research Method - Simulation Model and Simulation Environment

3.1. *Simulation Model*

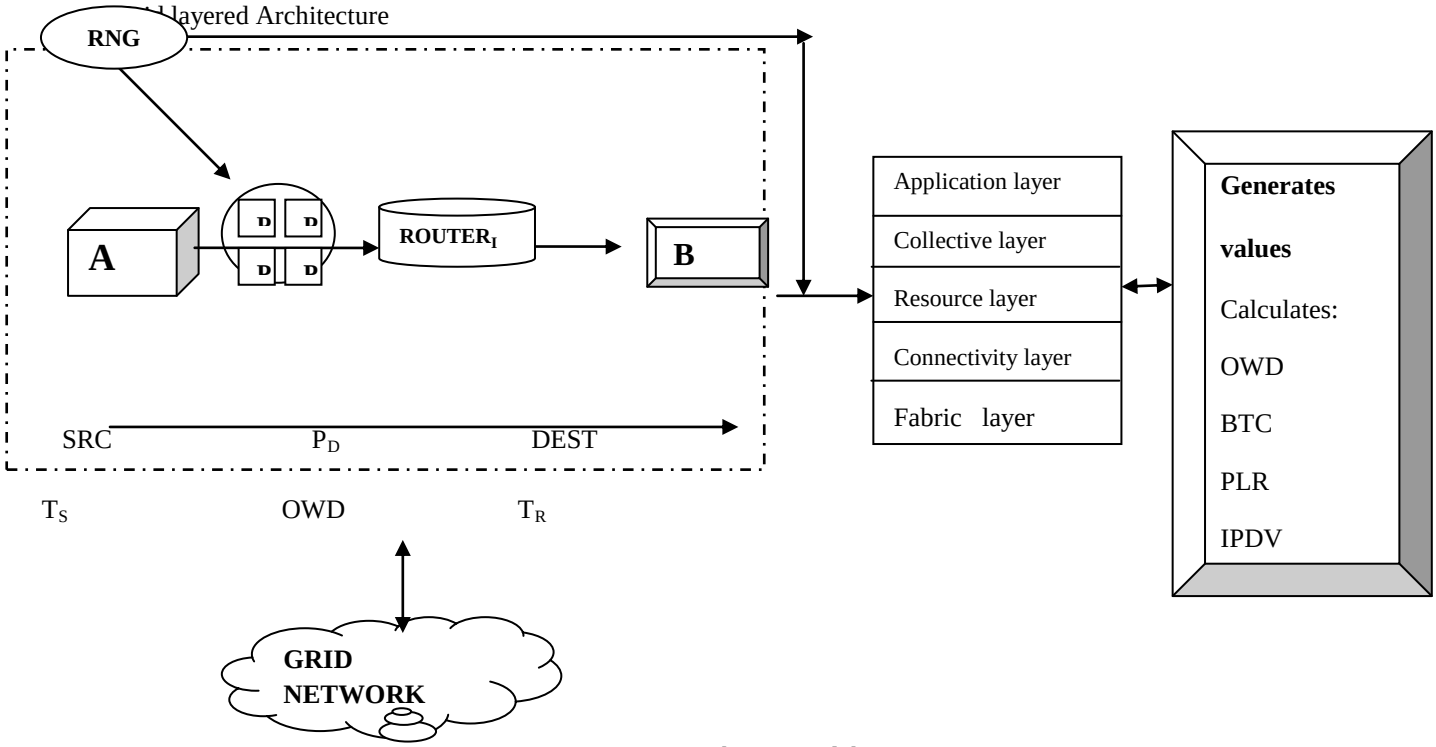

Figure 1: Simulation Model

Figure 1 is the model used to formalize the grid data transfer and represents the simulation of data transfer on the grid by first dividing the data into packets. Thereafter the model formalization is then translated into a simulation program. All necessary data transfers and synchronizations are performed by the GridFTP enabled servers.

Computer ***B*** which represents the (client system) identifies a data file in computer ***A*** that it needs. The sending computer ***A*** which represents a (server system), performs a check on the data file as to whether the data is available or idle, it then broadcasts the service level agreement (SLA) to computer ***B***, if the client system is satisfied with the service to be offered, then an acknowledgement is sent to computer ***A*** signally the initialization of data transfer.

The **gridftp** enabled server system subdivides data into smaller units called packets in order to ensure proper and reliable movement of the data along the network path. The two computers are connected with a **Router$_i$**.

Router$_i$ is an interconnectivity device that connects the client and server in order to facilitate the distribution of data. It also helps to find the correct path in which the packets should travel. For different computers to send data on the Grid, they must be connected to the grid network as shown in the figure 3. The paths in which the packets travel are indicated with the single arrows.

P_1, P_2, P_3 ……P_{100} are set of packets. Before data are been transferred over a network, it is first broken up into small units (packets) by the sending computer. The packets, which travel independently of one another, are marked with the sender's address, destination address, and other pertinent information, including data about any error introduced during the transfer.

When the packets arrive at the receiving computer, they are then reassembled into the original data and when the client has finished with the request data, the client forwards it back to the server system. For the purpose of this work, metrics are computed at the resource layer because that is the layer that guarantees access to data.

When a packet is sent along the network path, a random number generator (RNG) which is a process that produces random numbers generates information about the packets at the resource layer, the RNG afterwards records relevant attributes for the packets, after which the one-way delay metrics (OWD) is therefore

calculated, followed by the calculation of bulk transport capacity, packet loss ratio and IP delay variation metrics. Values for all the variables of the metrics are represented in real numbers. The results of the different simulation run for each of the metrics were therefore statistically analyzed.

Each packet $\mathbf{P_1, P_2, P_3 \ldots\ldots P_{100}}$ generated is characterized by the following properties:

$\mathbf{P_D}$: Time needed for a packet takes to go from one node to another on the network after which the router must have processed the packet.

$\mathbf{T_{Reception}}$: Time the packet takes to get to its destination (milliseconds).

$\mathbf{T_{Sent}}$: Time the packet was sent on the network from the source (milliseconds).

B: Number of bit transmitted from one grid node to the other.

Time-taken: Time it takes for the packet bit to move from one grid node to the other (seconds).

$\mathbf{OWD_i}$: Delay computation of the next packets i (i=1….100).

OWD: Delay computation of the current packets (milliseconds).

Packet sent: Number of packet sent from source in bits.

Packet received: Number of packet received at the destination in bits.

At the end of the process, a dataset was collected at to evaluate the metrics.

3.2 *Simulation Environment*

A computer simulation is a computer program that attempts to design an abstract model of a particular system. Simulations can be used to explore and gain new insights into new technology and to estimate the performance of systems too complex for analytical solutions [19]. It is a simpler yet effective approach for analyzing and evaluating designed mechanisms, protocols, algorithms or theories for systems.

Simulations provide methods to investigate newly developed protocols and their behaviour, performance and interaction with other protocols, validation, and feasibility, and to remove points of uncertainty. Random numbers are useful in simulating and modeling complex phenomena and for selecting random samples from larger datasets as it is on the Grid.

The simulation of the performance of Grid packets was carried out using MATLAB simulation software version 7.7. It was used to generate a set of random numbers for the packets and their relevant attributes, these values were then recorded and used to simulate each of the metrics. The results of different simulations were then statistically analyzed. This process of getting the results of an experiment from random number table is called simulation.

The MATLAB uses algorithm to generate pseudo-random numbers (PRN) in a specified distribution, these numbers use mathematical formulae. PRN [20] are efficient, that is, they can produce many numbers in a short period time, and deterministic, meaning that a given sequence of numbers can be reproduced.

Pseudo-random numbers in MATLAB come from one or more random number streams. The simplest way to generate arrays of random numbers is to use *rand*,*randn*, or *randi*. These functions all rely on the same stream of uniform random numbers, known as the default stream, other stream objects that act separately from the default stream can also be created, using *rand*, *randi*, or *randn* methods to generate arrays of random numbers. The sequence of numbers produced by rand is determined by the internal state of the uniform pseudo-random number generator that underlies *rand*, *randi*, and *randn* [21].

4. Results and Discussion

Figures 2, 3, 4, and 5 shows the graphical representations of the simulation results (QOS metrics) used to measure the performance of the data transfer using the developed simulation model (see figure 1).

4.1. *One-way Delay (OWD) Metrics*

This is the time that a packet takes to go from one point to another of the network, usually measured in milliseconds. It is further considered as the time elapsed since the first bit of the packet has been sent on the network until its last bit has reached its destination. The delay is specified from the start of the packet being transmitted at the source to the end of the packet being received at the destination. OWD is strictly positive real numbers. In "Eq. (1)", one-way delay metrics can be evaluated as:

$$\text{OWD} = \text{T reception}_{(i)} - \text{T sent}_{(i)} \quad (1)$$

Figure 2 which shows the simulation results of OWD indicates that, at a certain time the delay increases and at another time it decreases (an upward or downward movement) meaning that packets experience delay, although not constant but it varies over time i.e. delay is as a factor of the packet time of reception
at its destination. The delay could be due to network traffic which slows the processing time for each of the packets. When delay is below 35ms, it is a minimum delay and when greater than 35ms, it is a maximum delay. The study of this graph shows that for a typical data transfer on the grid, a delay of 84 ms of packets is expected to occur.

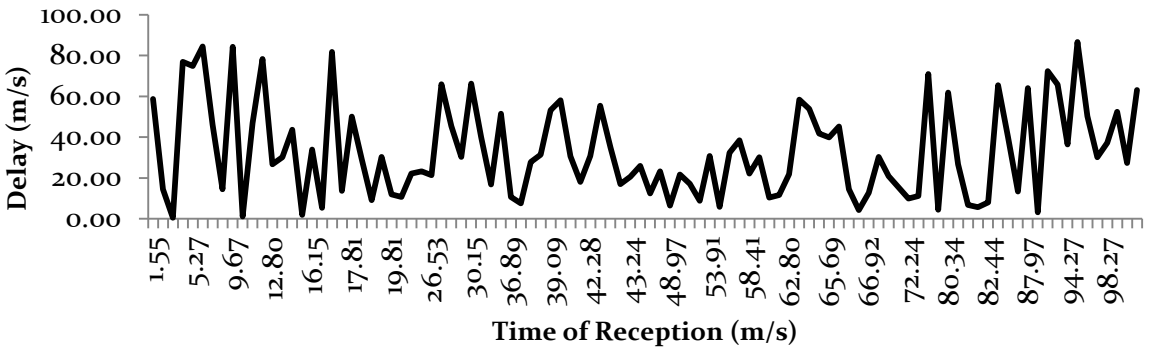

Figure 2: OWD versus Time of Reception

4.2. Bulk Transport Capacity (BTC) Metrics

Bulk Transport Capacity (BTC), is defined as the maximum transport capacity of a link using a single congestion-aware protocol connection (i.e., TCP). BTC metric presents of the maximum performance from a user point of view where big data transfers are involved (FTP, big HTTP downloads, etc.). It is a data transmission measure that determines the amount of packet bits moved from one node to another in a certain period of time, measured in megabits, kilobits or bits per second. In "Eq. (2)", bulk transport capacity can be evaluated as:

$$\text{BTC} = \frac{\text{Number of Bit Transmitted}}{\text{Time (seconds)}} \quad (2)$$

Figure 3 which shows the simulation results for Bulk Transport Capacity (BTC) metrics indicates that, as BTC decreases, the time in transmitting the packets increases. The apex of the graph shows the maximum BTC attained in the data transmission which is 1191Kbps at 0.019 seconds. The study of the graph shows that when the packets size in bits are large, it takes a shorter time to travel along the network path, thereby increasing the efficiency of data transmission, it takes a longer time to transfer data from one node to the other when the bulk transport capacity is below 119kbps.

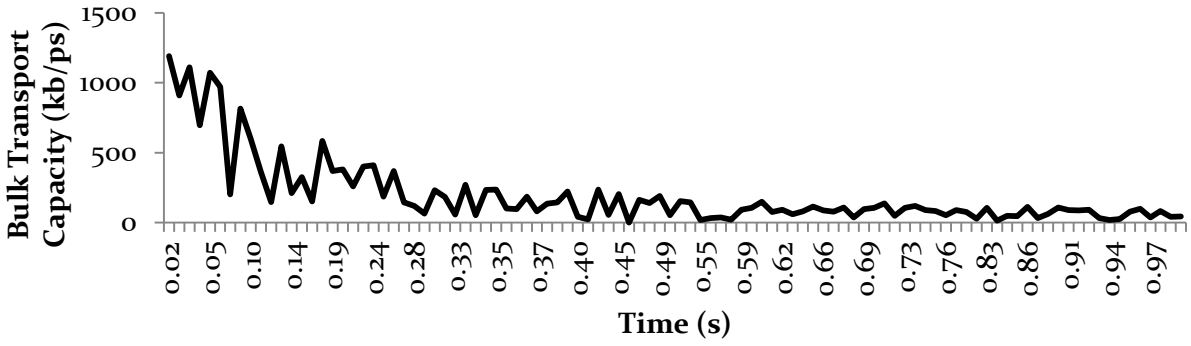

Figure 3: Bulk Transport Capacity (BTC) versus Time

4.3. Packet Loss Ratio (PLR) Metrics

This metric measures the success task in the transmission of packets between two nodes of the network, the smaller the loss of packet, the larger the efficiency of the network. Packet loss is an important performance

characteristic of network traffic, crucial for applications including long-range data transfers. The loss of packets may be due to packet errors indicating a bad link or packet drops as a result of congestion. In "Eq. (3)" below, the PLR metrics can be evaluated as:

$$\text{PLR} = \frac{\text{Packets sent} - \text{Packets received}}{\text{Packets sent}} \qquad (3)$$

Figure 4 presents the percentage of lost packets in data transmission. For the 100 generations of packets, there are variations to the number of packets received compared to the number of packets sent at the source. The packet loss ratio increased suddenly when the number of packet bits was 18, it decreases and increases continually due to the variation in the number of bits sent and received. The increase in packet loss ratio is sometimes due to the decrease in the number of packet received at the destination. The presences of this metrics however reduce quality and quantity of data received by the user. The network path to exhibit a packet loss ratio of 0.9% as the number of packets sent increases.

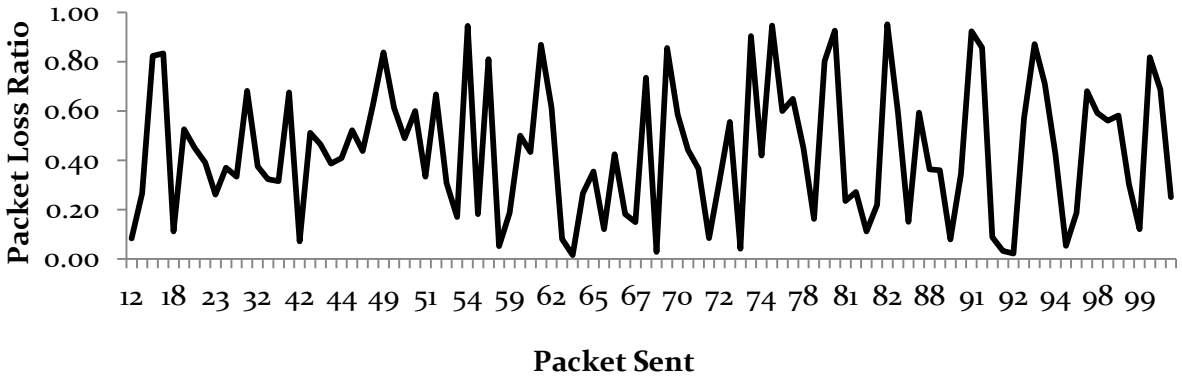

Figure 4: Packet Loss Ratio versus Packet Sent

4.4. Internet Protocol Packet Delay Variation (IPDV)

This is the differences between the delay of the current package and the next package also known as the successive packets. The successive sequences are not defined, but usually consecutive packets are selected. IPDV is a real number which could either be positive or negative. In "Eq. (4)", IPDV can be evaluated as:

$$\textbf{IPDV} = \textbf{OWD} - \textbf{OWD}_i \qquad \textbf{i} = 1, 2 \ldots\ldots 100 \qquad (4)$$

From figure 5, the IPDV shows variation in delay between sequential packets with some resulting in negativity. Because there are packet losses, IPDV values are therefore defined. The analysis of IPDV metrics allows us to observe whether the packets present a constant behaviour or there are lots of variations, the IPDV indicates an upward and downward direction.

Also, as seen in figure 5, the successive packets did not present a constant behaviour as IPDV varies along the network path, as the OWD in next packet increases IPDV values increases at certain time and decrease at another time. The maximum Delay variation of 81ms was therefore experienced.

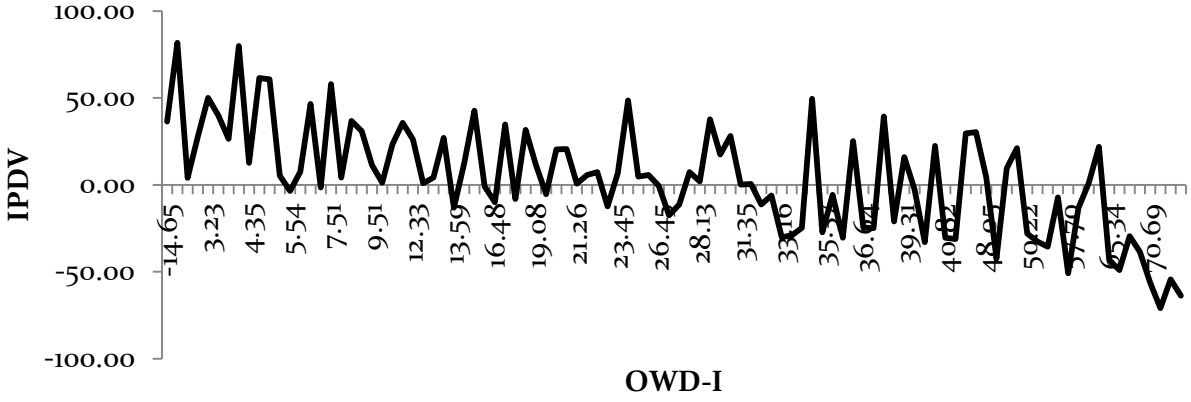

Figure 5: IPDV versus OWD of the next Packet

5. Summary and Concluding Remarks

This work basically describes the Quality of Service Metrics in Grid Computing focusing on the performance evaluation of network packets in the grid computing. It further includes a model to achieve the simulation which was used to analyze the behaviour and performance of the network packets using the one-way delay, bulk transport capacity, packet loss ratio and IPDV quality of service metrics. The application of graph was to have a general overview on the results. Simulation based evaluation method has been recognized as the best way for performance evaluation in research and development. The need to subdivide data into packet in order to guarantee that a reliable and complete data are delivered to the user has been emphasized. Due to the required high volume of data transfer on the grid, random samplings of packets were used as a mechanism to reduce the dataset. Data set were statistically analyzed to summarize the results from the experiments and to look-up for desirable properties.

The simulation results showed the following that:

i. The higher the delay encountered, the lower the bulk transport capacity.
ii. Packets experience delay which are not constant but rather changes over time.
iii. The proportion of the packet sent to that of the packets received will determine the amount of packet loss experienced in a data transmission.
iv. A change in any variable of the metrics will affect the quality of data transferred.
v. Due to the packets loss ratio, IPDV values are defined.
vi. Time is very essential in data transmission to guarantee quality of service.

The use of QOS metrics is therefore recommended for evaluation of the performance of network packet in transfer of data in the grid environment. Future work may to optimized data access in grid computing.

References

[1]. Akinyemi, I. O.; Daramola, J. O.; Adebiyi, A. (2007): A: Grid-Enabled e-Learning Framework for Nigerian Educational Institutions, Proceedings of the Annual Conference of Nigeria Computer Society, MILDEG 2007, pp. 91-98.

[2]. Foster, I.; Kesslman, C. (2004): The Grid Blueprint for a new Computing Infrastructure, 2nd edition, San Franscisco, Morgan Kaufmann publishers, retrieved June 19, 2011 from http://www.mkp.com/books catalog/1-55860-475-8.asp.

[3]. Foster, I.; Kesselman, C. (1999): The Grid: Blueprint for a New Computing Infrastructure, San Francisco, CA: Morgan Kaufmann publishers, p. 7.

[4]. Gullapalli, S.; Czajkowski, K.; Kesselman, C; Fitzgerald, S. (2011): The grid notification framework, Technical Report, Grid Forum Working Draft GWD-GIS-019, retrieved June 19, 2011 from http://www.gridforum.org.

[5]. Tuecke, S.; Czajkowski, K.; Foster, I.; Frey, J.; Graham, S.; Kesselman, C. (2002): Grid Service Specification, retrieved May 27, 2011 from http://www.globus.org

[6]. Berman, F.; Fox, G.; Hey, T, (2003): Grid Computing: Making the Global Infrastructure a Reality, John Wiley & Sons Ltd, p. 9.

[7]. Thain, D.; Tannenbaum, T. ; Livny, M. (2003): Condor and the grid in Grid Computing: Making the Global Infrastructure a Reality, Chichester, John Wiley & Sons, Ltd, p. 300.

[8]. Serral-Gracià, R.; Domingo-Pascual, J.; Beben, A.; Owezarski, P. (2008): QoS Measurements in IP-based Networks in End-to-End Quality of Service Over Heterogeneous Networks, Springer-Verlag Berlin Heidelberg, pp. 22-28.

[9]. Brandley, M. (2011): Computer Networking Definition for Packet, retrieved May 26, 2014 from http://compnetworking.about.com/od/networkprotocols/l/bldef_packet.htm.

[10]. Gokhale, A. (2005): Introduction to Telecommunications, 2nd edition, Thomason Delmar Learning, p. 252.

[11]. Marchese, M. (2007): QoS Over Heterogeneous Networks, John Wiley and Sons Ltd, p. 212.

[12]. Braun,T.; Staub, T (2008): Motivation and Basics of QOS Parameters in End-to-End Quality of Service Over Heterogeneous Networks, Springer-Verlag, Berlin Heidelberg, pp. 2 - 4.

[13]. Foster, I. (2002): What is the Grid: A Three Point Checklist,*GRID Today*, **1**(6), available at http://www.fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf.
[14]. Allcock, B.; Bester, J., Bresnahan, J.; Chervenak, L.; Foster, I.; Kesselman, C.; Meder, S.; Nefedova, V.; Quesnel, D.; Tuecke, S. (2002): Data Management and Transfer in High Performance Computational Grid Environments", *Parallel Computing*, **28**(5), 749–771.
[15]. Setty, S.; Narasimha, S.; Rraju, K.; Naresh, K. (2010): Performance Evaluation of AODV in different Environments, CS&SE Department, Andhra University College of Engineering, India.
[16]. Colling, D.; Ferrari, T.; Hassoun, Y.; Huang, C.; Kotsokalis, C.; Mcgough, A.; Patel, Y.; Tsanakas, P. (2002): On Quality of Service Support for Grid Computing, *Collections*, London, chapter 1, p. 2.
[17]. Menasce, D.; Casalicchio, E. (2014): Quality of Service Aspects and Metrics in Grid Computing, Proc. 2004 Computer Measurement Group Conference, Las Vegas.
[18]. Kalogeraki, V.; Melliar-Smith, P.; Moser, L. (2000): Dynamic scheduling of distributed method invocations, Proceedings of The 21st IEEE Real-Time Systems Symposium, pp. 57-66.
[19]. Strogatz, S. (2007): The End of Insight in What is your dangerous idea?, HarperCollins Ltd, New York.
[20]. Wetzel, C. (2007): Introduction to Randomness and Random Numbers, retrieved May 26, 2011 from www.random.org.
[21]. Mathworks International (2008), retrieved June 27, 2011 from www.mathwork.com