Performance Prediction of a Weighted Capacity Sharing Scheme for Grid Bulk Data Transfers using a Multi-service Queue

Israr Ullah[1], Kashif Munir[2]
National University of Computer and Emerging Sciences (NUCES)
Islamabad, Pakistan.
[1]israrullahkk@yahoo.com, [2]kashif.munir@nu.edu.pk

***Abstract*—Applications in high-speed networks like Grids often require to transfer large volumes of data to remote locations under certain time constraints. This requires a network reservation, mechanism for the large data transfer requests. It is possible that some network capacity is available even after fulfilling the capacity requirements of the requests. The sharing of the available network capacity, called residual capacity, has direct impact on the actual time taken to serve a request. In this paper, we propose a novel weighted capacity sharing model to predict the performance of the deadline-constrained data transfers. A weighted sharing scheme is desirable since a request with higher capacity requirement should get higher share of the residual capacity than that of a request with lower capacity requirement in order to reduce its transfer time. We calculate the blocking probabilities and mean flow time using our scheme and compare the results with an existing scheme of equal residual capacity sharing. The results show that weighted-sharing scheme is better than equal capacity-sharing scheme as it reduces the mean transfer time of requests of a higher capacity requirement class. We also validate the model through simulations.**

I. Introduction

Grid computing is a combination of loosely coupled, heterogeneous computer resources which are geographically dispersed across multiple administrative domains. Although a Grid can be dedicated for a specific application, it is more common that a single Grid is used for a variety of different purposes like solving computationally intensive scientific and mathematical problems, drug discovery, economic forecasting, seismic analysis, and back office data processing in support for e-commerce and web services etc.

Since the start of 21st century, many Grid network reservation schemes like [1], [2], [3], [4], [5] have been proposed. The schemes target the scheduling of Grid bulk data transfer requests within certain time bounds. A request whose capacity requirement can be fulfilled from its start time till its deadline is admitted in the network and the network resources are reserved for it. In the other case, the request is rejected. It is possible to revise the deadline of the rejected request and hence its capacity requirement for its re-submission. At any particular moment, it is possible that some network capacity is available even after fulfilling the capacity requirements of the reserved requests. Onwards, we call this available network capacity as residual capacity.

A survey on statistical bandwidth sharing for elastic traffic in computer networks is given in [6]. In [6] two types of bandwidth sharing are described. First type is the fair sharing and the second is the unfair sharing also called as discriminatory sharing. Max-min fairness is a type of fair sharing in which the rate of individual flows is made as equal as possible. This concept was used in [7], where the residual capacity is equally shared among active flows. According to a type of unfair sharing mentioned in [6], the share attributed to a user of class K is proportional to a W_K. We use this concept to share the residual capacity among the active flows and call it Weighted-Sharing. We determine the weight of a class on the basis of the capacity requirement of the flows belonging to that class.

We are interested in developing an analytical model for multi-class deadline-constrained data transfers in high speed network. The model will be a representation of a system with multi-class users, each having its own deadline constraints imposed on their data transfer requests. Example of such a system can easily be found in Grid/Cloud computing. The model is only applicable to high speed networks because the notion of deadline-constrained data transfers is only valid in such networks while traditional Internet is known for its best effort services.

In this paper, we use a multi-service queue to predict the performance of deadline-constrained Grid bulk data transfer requests having weighted sharing of residual capacity. This work is an extension of the work presented in [7], where the residual capacity is equally shared among the active flows. We categorize the data transfer requests in different traffic classes based on their capacity requirements. The weighted-sharing scheme favors the request of a higher capacity requirement class by reducing its mean transfer time than the request of a lower capacity requirement class.

The rest of the paper is organized as follows: the related work is presented in Section II. Problem formulation is done in Section III. We discuss the analytical model in Section IV. Performance evaluation results are discussed in Section V. We conclude the paper in Section VI.

II. Related Work

Network flows are elastic in nature as their rates can be adjusted by opportunistically grabbing the available bandwidth

978-1-4577-0768-1/11/$26.00 ©2011 IEEE

beyond their current demands. Modeling elastic traffic is relatively a new field but has received rapid attention in recent past. We briefly describe the literature that is closely related to our work.

Bonald et al. has worked on performance modeling of elastic traffic in [8] without considering deadline constraints of data transfers which is a key difference from our work. To approximate the mean throughput of TCP, the authors have modeled the fair sharing bottleneck with an M/G/1-PS queue. To calculate the mean probability of per connection share in high speed networks, Berger et al. have proposed a model in [9] for bandwidth dimensioning, considering a single bottleneck link. The authors of [10] have used M/M(a,b)/c/PR priority queue to model the semi-conductor manufacturing operations, considering only two priority classes. To analyze the control schemes for 3G wireless networks, AlQahtani et al. have proposed a model in [11]. They have analyzed two real-time and two non-real-time traffic classes. Similarly, Fodor G. et al. have work on calculation of blocking probabilities and throughput guarantees in [12] for three different classes of flows.

The proposed model can be used for any number of classes having different capacity requirements. All classes are given same priority in this work. This work is an extension of the work presented in [7] where the residual capacity is equally shared among the flows. In this paper, we have presented a novel model to enable weighted sharing of residual capacity. It favors the flows of a higher capacity class by reducing their mean flow time as they are assigned more share from the residual capacity without affecting the blocking probability of individual classes.

III. Problem Formulation

Definitions:

1) Data Transfer Request: A data transfer request $r = (\nu_r, \omega_r, \phi_r)$ is a tuple, where ν_r is the volume of r, $\omega_r = [\eta_r, \psi_r]$ is the active window (from arrival time η_r to deadline ψ_r) and ϕ_r is the path connecting source S_r and destination D_r of the request r.
2) MRR_r: MRR_r is the Minimum Required Rate of the request r. It is calculated on the basis of its volume and active window as follows:
$$MRR_r = \frac{\nu_r}{\psi_r - \eta_r}$$
3) BP: Blocking Probability is the ratio of total rejected requests and total number of submitted requests.
4) MFT: Mean Flow Time is the average time to serve a request.

Consider a shared bottleneck link having capacity C. Data transfer requests are categorized into R classes on the basis of their minimum required rates. A request is accepted if its MRR_r can be fulfilled. At any time instant t, a request of an ith class is accepted if

$$C_r \geq MRR_r$$

where C_r is the residual capacity of the link and can be calculated as follows:

$$C_r = C - \sum_{i=1}^{R} MRR_i \times N_i \quad (1)$$

where N_i is the number of requests of ith class.

The state of the system S at any time instant t, can be represented as:

$$S = (N_1, N_2, N_3, \ldots, N_R)$$

There can be three possibilities if the residual capacity $C_r \geq 0$. We use the terms request and flow interchangeably.

- No-Sharing (NS): C_r is not shared among the active flows in this scheme. It results in under-utilization of network capacity. We have only considered this scheme for comparison purpose.
- Equal-Sharing (ES): It is equal sharing of C_r among active flows irrespective of which class a flow belongs to. In [7], sharing of C_r is done in this way. This scheme is better than No-sharing.
- Weighted-Sharing (WS): It is the sharing of C_r on the basis of the MRR of active flows. In this scheme, more share of C_r is given to a request of a higher MRR class than it is given to a request of a lower MRR class. The advantage of this scheme over the scheme presented in [7] is reduction in the mean transfer time of a request of a higher MRR class which can also be considered as a higher priority class.

Suppose that C is 15 $Gbps$ and $R = 3$. Let's assume that the system is in state $S = (1, 2, 1)$ i.e. there is one flow of class 1, two flows of class 2 and one flow of class 3. Thus the total occupied capacity of link is 8 $Gbps$ and residual capacity C_r is 7 $Gbps$. Figure 1 shows the sharing of C_r according to the three schemes mentioned above.

One of the objectives of the study is to find BP and MFT. BP is a critical indicator of a system's behavior and performance and can be used in dimensioning of the network [7]. BP is not linearly dependent on the link capacity or on traffic intensity and, therefore, cannot be directly calculated. In the next section, we describe our model to calculate BP and MFT of overall system as well as of an individual class.

IV. M/M/1/K-WS Model

The bottleneck link is considered as a constant capacity transmission server which services multi-class data transfer requests. Arrivals of requests in Grid networks do not follow any known distribution, but due to large number of requests, arrivals can be considered as Poisson. For the sake of simplicity, without any loss of generality, we assume that volume of the requests of each class is Exponentially distributed with mean volume V. K is number of requests in the queue depending on the current state of the system.

We compute BP and MFT by modeling the system as an R-dimensional Continuous Time Markov Chain (CTMC) as shown in Figure 2. A state S of the system is uniquely

978-1-4577-0768-1/11/$26.00 ©2011 IEEE

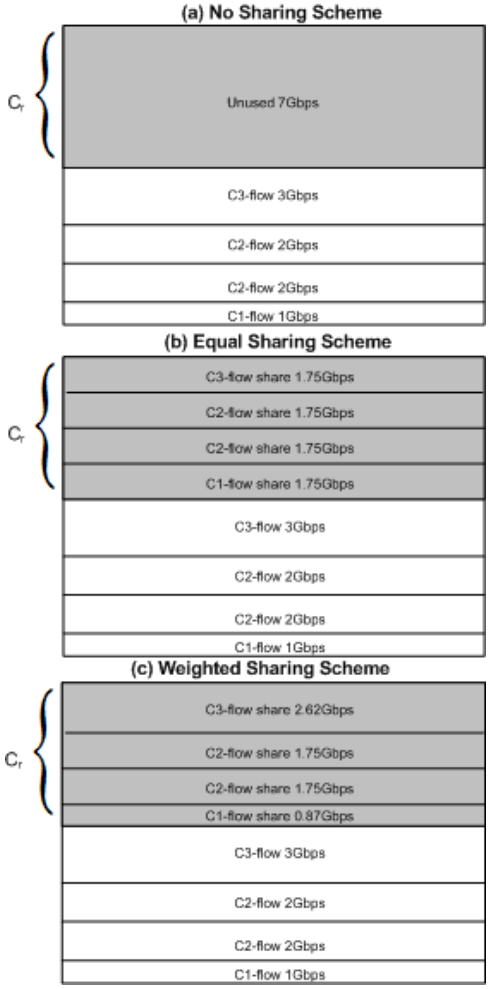

Fig. 1. Sharing of Residual Capacity according to various schemes.

identified by the number of existing requests of each class i.e. $S = (N_1, N_2, N_3, \ldots, N_R)$. Arrivals and departures of requests cause transitions among different states of a Markov chain.

Given the system states

$$S_i = (N_1, N_2, \ldots, N_c, \ldots, N_R)$$

$$S_j = (N_1, N_2, \ldots, N_c + 1, \ldots, N_R)$$

$$S_k = (N_1, N_2, \ldots, N_c - 1, \ldots, N_R)$$

As requests of every cth class is generated according to Poisson process with parameter λ_c, the rate of transition from state S_i to S_j is given as follows:

$$\lambda_{i,j} = \lambda_c \tag{2}$$

Let the system moves from state i to state k upon completion of a flow of class c. This transition will occur with different service rates for each of the three schemes mentioned in Section III. Using NS scheme, the service rate is calculated as follows:

$$\mu_{i,k} = \begin{Bmatrix} \frac{N_c}{V}(MRR_c) & forN_c > 0 \\ 0 & otherwise \end{Bmatrix} \tag{3}$$

where V is the mean volume of the data transfer requests. For ES scheme, the service rate is calculated as follows:

$$\mu_{i,k} = \begin{Bmatrix} \frac{N_c}{V}\left(MRR_c + \frac{C_r}{\sum_{i=1}^{R} N_i}\right) & forN_c > 0 \\ 0 & otherwise \end{Bmatrix} \tag{4}$$

For WS scheme, the service rate is calculated as follows:

$$\mu_{i,k} = \begin{Bmatrix} \frac{N_c}{V}\left(MRR_c + \frac{C_r \times MRR_c}{\sum_{i=1}^{R} N_i \times MRR_i}\right) & forN_c > 0 \\ 0 & otherwise \end{Bmatrix} \tag{5}$$

Figure 2 shows a sample CTMC with R=3, C=4 Gbps.

Solving the CTMC will give us the steady state probability vector π of the system in being all valid states S. If the total numbers of valid states in CTMC are n then

$$\pi = (p_1, p_2, p_3, \ldots, p_n)$$

where p_i is the probability of the system being in state i. The method used for calculating π vector is explained below:

A state S is valid if:

$$\sum_{i=1}^{R} (N_i \times MRR_i) \leq C$$

Let Θ be the set of all valid states. We generate all valid states and compute the infinitesimal generator matrix Q using Equation 2 for calculation of arrival rates, Equations 3, 4 and 5 for service rates. Solving the CTMC with n states corresponds to solving the set of steady-state equations of the form:

$$\pi Q = 0$$

under the constraint $\sum_{i=1}^{n} p_i = 1$.

In order to transform infinitesimal generator matrix Q into one step transition probability matrix P, first, we find the maximum absolute value $MaxDiag$ at the diagonal of Q i.e.

$$MaxDiag = max(|Q_{x,y}|) \ \forall \ x = y$$

Then, we obtain Q' matrix by dividing every element of Q by MaxDiag as follows:

$$Q' = \frac{Q_{x,y}}{MaxDiag} \ \forall \ x, y = 1, 2, 3, \ldots, n$$

Finally, we obtain one step transition probabilities matrix P as follows:

$$P_{x,y} = \begin{Bmatrix} Q_{x,y} & forx \neq y \\ Q_{x,y} + 1 & forx = y \end{Bmatrix}$$

Once, we obtain the one step transition probability matrix P, the Iterative (Power) method [13] can be used to calculate the steady state probability vector π as follows:

$$\pi^0 P = \pi^1$$

978-1-4577-0768-1/11/$26.00 ©2011 IEEE

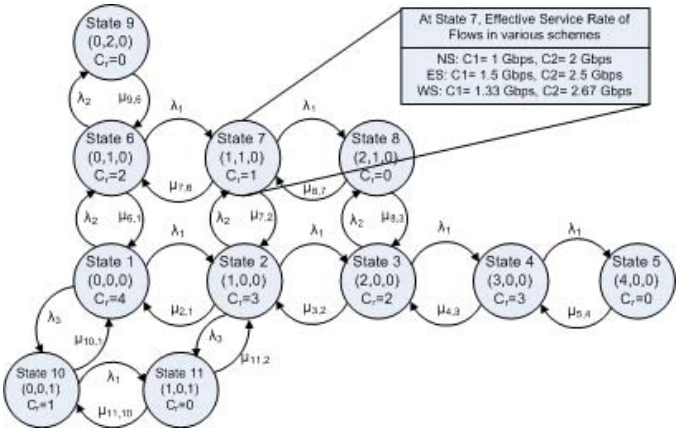

Fig. 2. Sample CTMC with link capacity C=4 Gbps and Number of classes R=3 and $MRR_c = c\ Gbps\ \forall\ c \in \{1, 2, \ldots, R\}$

$$\pi^1 P = \pi^2$$

$$\ldots$$

$$lim_{n\to\infty} \pi^n P = \pi^n$$

where π^0 is initial (random) probability distribution vector with condition $\sum_{i=1}^{n} p_i = 1$.

A. Computation of BP

This steady state probability vector

$$\pi = (p_1, p_2, \ldots, p_s, \ldots, p_n)$$

is then used to calculate of blocking probabilities of the system and of each class, where p_s is the probability of the system being in state $s \in \{1, 2, 3, \ldots, n\}$. In order to calculate BP of a particular class c, we need to identify all those states in which a new request of cth class can not be accommodated. Let S_{B_c} be the set of all those states. Blocking probability of cth class BP_c is calculated as follows:

$$BP_c = \sum_{\forall s \in S_{B_c}} p_s$$

The overall blocking probability of the system is calculated as follows:

$$BP = \left[\sum_{c=1}^{R} \left(\sum_{\forall s \in S_{B_c}} p_s \times \lambda_c \right) \right] \times \frac{1}{\lambda}$$

B. Computation of Mean Flow Time

Mean number of flows L in the system is calculated from steady state probability vector π as follows:

$$L = \sum_{\forall s \in \Theta} p_s \times (N_1 + N_2 + \cdots + N_R)$$

To find mean number of flows L_i of ith class, we have:

$$L_i = \sum_{\forall s \in \Theta} p_s \times (N_i)$$

Using the mean number of flows of ith class L_i, we can calculate the mean reserved capacity C' as follows:

$$C'_{used} = \sum_{i=1}^{R} L_i \times (MRR_i)$$

Hence, mean residual capacity C'_r is given by

$$C'_r = C - C'_{used}$$

In the case of ES scheme, the effective transfer rate of a flow of an ith class is given below:

$$(\overline{MRR_i})_{ES} = MRR_i + C'_r \times \frac{1}{L}$$

Whereas in the case of WS scheme, it becomes:

$$(\overline{MRR_i})_{WS} = MRR_i + C'_r \times \frac{MRR_i}{C'_{used}}$$

Finally, MFT of a flow for all three schemes is calculated as follows:

$$MFT_{NS} = \frac{V}{MRR_i}$$

$$MFT_{ES} = \frac{V}{(\overline{MRR_i})_{ES}}$$

$$MFT_{ES} = \frac{V}{(\overline{MRR_i})_{WS}}$$

V. Performance Evaluation

The objectives of the performance evaluation are:

- Validation of the model.
- Comparison of BP and MFT of the schemes mentioned in Section III.

The model is validated using an adhoc simulator implemented in $VB.NET$. The simulations are performed to validate the model and hence we do not consider the network level overheads and losses in the simulations. For each simulation, we generate 100,000 requests according to Poisson process. We generate the volumes of requests according to Exponential distribution with mean V. Table I shows the values of various

978-1-4577-0768-1/11/$26.00 ©2011 IEEE

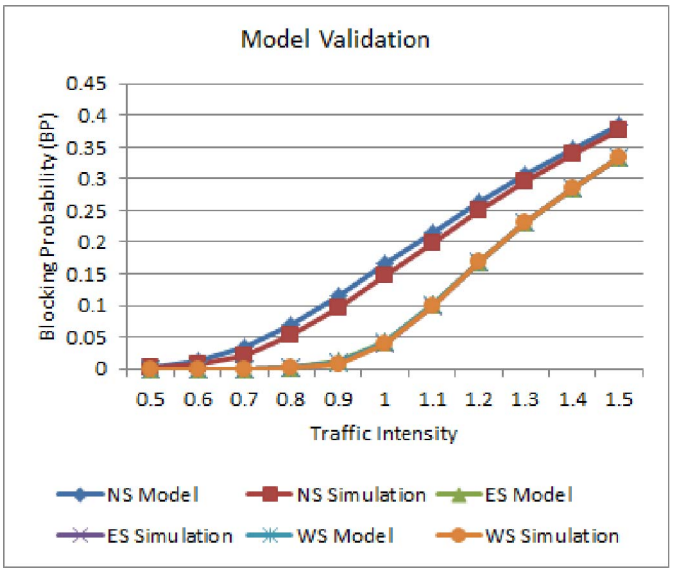

Fig. 3. Model validation against Simulation for $C = 40$ $Gbps$, No. of classes R=3 and $MRR_c = c$ $Gbps$ $\forall$ $c \in \{1, 2, \ldots, R\}$

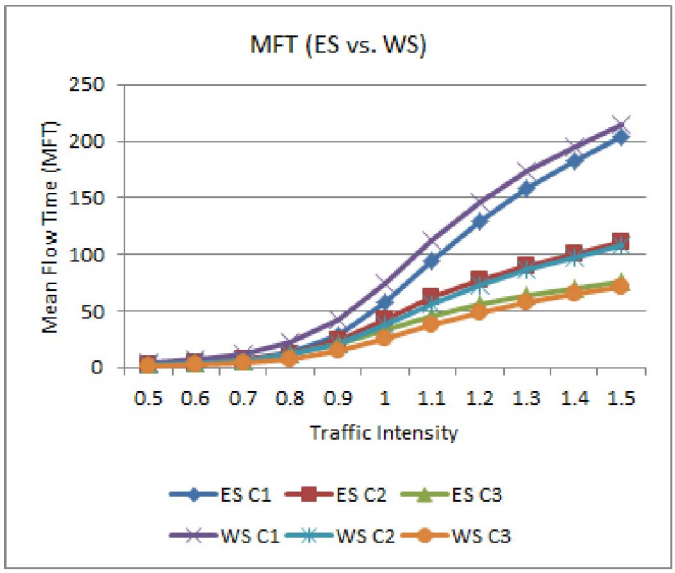

Fig. 4. Mean Flow Time results for $C = 40$ $Gbps$, No. of classes R=3 and $MRR_c = c$ $Gbps$ $\forall$ $c \in \{1, 2, \ldots, R\}$

parameters used in the simulations. Each simulation is run 10 times for same parameters, and average values are reported.

For the sake of simplicity, we assume that $MRR_c = c$, $\forall$ $c \in \{1, 2, \ldots, R\}$ and all classes' arrivals are equally likely. Thus, the arrival rate of the request of a particular class c is given by:

$$\lambda_c = \frac{\lambda}{R}$$

Traffic intensity ρ is calculated as follows:

$$\rho = \frac{\lambda \times V}{C}$$

TABLE I
VALUES OF PARAMETERS USED IN SIMULATION AND MODEL.

S/No.	Parameter	Value/Range	
		Model	Simulation
1	Arrival rate	0.25	0.25 (exp. distributed)
2	Traffic intensity	0.5 - 1.5 (step 0.1)	0.5 - 1.5 (step 0.1)
3	Link Capacity	40 Gbps	40 Gbps
4	Flow size	80 - 240 (step 16)	80 - 240 (step 16, exp. distributed)
5	No. of classes	3	3
6	Number of Requests	NA	100,000

Figure 3 shows the blocking probabilities which are calculated for various traffic intensities using the analytical model and simulation. The results show that the simulations validate the model. The BP is significantly higher in NS scheme as compared to the BPs in other two schemes which is obvious as C_r is wasted in the case of NS.

BP is almost same for ES and WS schemes as shown in the Figure 3. This is due to the reason that C_r is not wasted in either case. However, MFT is reduced for higher MRR classes and increased for lower MRR classes if WS is compared with ES as shown in Figure 4. The reason is obvious because higher class flows are given relatively more share of residual capacity and hence processed in shorter period of time. Similarly, lower class flows gets small share of residual capacity and their MFT is increased.

Table II shows the same results in a tabular form with individual class MFT of both ES and WS schemes. The last column shows the percentage difference between respective classes MFT. Up and Down arrows along with values show the percentage increase and decrease in MFT of each class respectively. In WS scheme, the MFT of higher class requests is decreased without causing any effect on the BP. For R=3 and C=40 Gbps, results shows that $MFTs$ of $C2$ and $C3$ requests are reduced by 35.58% and 11.23% respectively, while MFT of $C1$ is increased by 83.84% at maximum.

TABLE II
PERCENT INCREASE/DECREASE IN MFT OF INDIVIDUAL CLASSES FOR ES AND WS SCHEMES.

ρ	ES			WS			Percent Increase/Decrease		
	C1	C2	C3	C1	C2	C3	C1	C2	C3
0.5	2.05	2.00	1.95	3.77	1.88	1.26	⇑ 83.84	⇓ 5.72	⇓ 35.58
0.6	3.73	3.59	3.46	6.66	3.33	2.22	⇑ 78.38	⇓ 7.34	⇓ 35.91
0.7	6.88	6.49	6.13	11.79	5.90	3.93	⇑ 71.29	⇓ 9.09	⇓ 35.88
0.8	13.52	12.23	11.16	21.79	10.90	7.26	⇑ 61.19	⇓ 10.89	⇓ 34.92
0.9	28.54	23.82	20.44	41.78	20.89	13.93	⇑ 46.41	⇓ 12.29	⇓ 31.85
1	57.05	42.05	33.30	74.66	37.33	24.89	⇑ 30.88	⇓ 11.23	⇓ 25.26
1.1	94.52	61.49	45.57	113.01	56.50	37.67	⇑ 19.56	⇓ 8.12	⇓ 17.34
1.2	129.88	77.47	55.20	146.63	73.31	48.88	⇑ 12.89	⇓ 5.37	⇓ 11.46
1.3	158.72	90.02	62.83	173.22	86.61	57.74	⇑ 9.14	⇓ 3.79	⇓ 8.10
1.4	182.62	100.60	69.42	195.23	97.61	65.08	⇑ 6.90	⇓ 2.97	⇓ 6.26
1.5	203.71	110.18	75.52	214.87	107.43	71.62	⇑ 5.48	⇓ 2.50	⇓ 5.16

VI. CONCLUSIONS AND FUTURE WORK

In a multi-service environment, it is desirable that flows of a higher capacity class should be given more share of residual capacity than the flows of a lower capacity class. In this paper, we have presented a novel model for multi-class deadline-constrained Grid bulk data transfers using weighted sharing

978-1-4577-0768-1/11/$26.00 ©2011 IEEE

of residual capacity. WS scheme allows the flows with higher MRR to get more share of the residual capacity which results in reduction of their MFT. We have calculated BP and MFT results for both, ES and WS, schemes. The results show that the weighted sharing of the residual capacity is better than equal sharing as it favors the request of a higher capacity class in terms of reduction in MFT while the BP of individual classes stays the same as it is in ES scheme.

The work can be extended to a multi-service priority based system where priorities of classes can be defined on the basis of their capacity requirements i.e. modeling of a system in which a request belonging to a higher capacity class can be given pre-emptive priority over a request belonging to a lower capacity class.

REFERENCES

[1] F. Bouabache, T. Herault, S. Peyronnet, F. Cappello, *Planning Large Data Transfers in Institutional Grids,* IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010.

[2] K. Rajah, S. Ranka, and Y. Xia, *Advance Reservation and Scheduling for Bulk Transfers in Research Networks,* IEEE Transactions on Parallel and Distributed Systems, 2009, 20(11):1682-1697.

[3] S. Figueira, N. Kaushik, S. Naiksatam, S. A. Chiappari, and N. Bhatnagar, *Advance Reservation of Lightpaths in Optical Network Based Grids,* ICST/IEEE GridNets, 2004.

[4] K. Munir, S. Javed, and M. Welzl, *A Reliable and Realistic Approach of Advance Network Reservations with Guaranteed Completion Time for Bulk Data Transfers in Grids,* Proc. ACM International Conference on Networks for Grid Applications (GridNets '07), Oct. 2007.

[5] S. Soudan, B.B. Chen, P. Primet Vicat-Blanc, *Flow scheduling and end-point rate control in GridNetworks,* Elsevier Future Generation Computer Systems, 2009, 25(8):904-911.

[6] J. Roberts, *A survey on statistical bandwidth sharing,* Elsevier Computer Networks, 2004, 45(3):319-332.

[7] K. Munir, P. Primet Vicat-Blanc, M. Welzl, *Grid Network Dimensioning by Modeling the Deadline Constrained Bulk Data Transfers,* 11th IEEE International Conference on High Performance Computing and Communications (HPCC 2009), Seoul, South Korea, 2009.

[8] T. Bonald, and J. Roberts, *Performance modeling of elastic traffic in overload*, Proceedings of the 2001 ACM SIGMETRICS international conference on measurement and modeling of computer systems, 2001, pp. 342-343.

[9] Berger A.W., Kogan Y., *Dimensioning Bandwidth for Elastic Traffic in High-Speed Data Networks,* IEEE/ACM Transactions on Networking, 2000; 8(8):643-654.

[10] Phojanamongkolkij N., Cochran J.K., Fowler J.W., *Multi-Products Multi-Servers Bulk Service Queue with Threshold Service Size,* In Proc. International Conference on Semiconductor Manufacturing Operational Modeling and Simulation, 1999, 153-156.

[11] AlQahtani S.A., Mahmoud A.S., *Performance analysis of two throughput-based call admission control schemes for 3G WCDMA wireless networks supporting multiservices,* Computer Communications, 2008; 31(1):49-57, Elsevier.

[12] Fodor G., Racz S., Telek M., *On Providing Blocking Probability- and Throughput Guarantees in a Multi-service Environment,* International Journal of Communication Systems, 2002; 15(4):257-285.

[13] William J. Stewart, *Probability, Markov Chains, Queues, and Simulation*, Princeton University Press, Princeton, NJ., QA273.S7532: 2009. ISBN 0-691-03699-3.

978-1-4577-0768-1/11/$26.00 ©2011 IEEE